

**MONITORING, VERIFICATION AND EVALUATION UNIT
AGRICULTURAL POLICY REFORM PROGRAM**

**MVE UNIT
APRP**

Sponsored by:

**Government of Egypt,
Ministry of Agriculture and Land Reclamation**

**United States Agency for International Development/Cairo
Office of Economic Growth, Agricultural Policy Division**

**AVAILABILITY
AND QUALITY
OF
AGRICULTURAL
DATA IN EGYPT**



Abt Associates Inc.

Prime Contractor:
Abt Associates Inc.

Subcontractors:
**Environmental Quality International,
Management Systems International**

USAID Award: 263-C-00-97-00003-00

Project Office: 15th Floor, 7 Nadi El Seid Street, Dokki, Cairo
Telephones: (202) 337-0357, 337-0592, 337-0378
Fax: (202) 349-9278

Morsy A. Fawzy
EQI

Larry Pope
Abt Associates Inc.

Yehia M. El-Din
Ghazal Amer
EQI

December 1998

Impact Assessment
Report No. 4

TABLE OF CONTENTS

LIST OF TABLES	iv
LIST OF CHARTS	vi
LIST OF ACRONYMS	vii
ACKNOWLEDGMENTS	viii
EXECUTIVE SUMMARY	ix
1. INTRODUCTION	1
1.1 Agricultural Policy and Statistics	1
1.2 Organizational Structure of Agricultural Data Collection and Publications ..	3
1.3 Objectives of the Study	3
1.4 Outline of the Study	4
2. CRITERIA FOR DATA QUALITY AND AVAILABILITY	5
2.1 Types of Data	5
2.1.1 Time Dimension and Intended use	5
2.1.2 Quantitative vs. Qualitative Data	6
2.2 Types of Errors	6
2.2.1 Non-Sampling Errors	6
2.2.2 Sampling Errors	6
2.2 Non-Response Errors	7
2.3 Criteria For Evaluating Information	7
2.3.1 Timeliness	7
2.3.2 Completeness of Coverage	7
2.3.3 Accuracy	8
2.4 Evaluating Data Quality	8
2.5 Methods to Improve Data Quality	9
2.5.1 Quality Control	9
2.5.2 Check Data	9
3. COLLECTION OF AGRICULTURAL DATA IN EGYPT	10
3.1 Structure of Agricultural Data Collection Organizations	10
3.2 Description of Data Collection and Estimation Procedures	13
3.2.1 Agricultural Census, EAS/MALR	13
3.2.2 Current Agricultural Statistics, EAS/MALR	13
3.2.3 Egyptian Survey Authority, MPWWR	14
3.2.4 Sampling General Directorate	15
4. METHODOLOGY	20
4.1 Investigation of Statistical Offices	20
4.1.1 Identifying Data Available at Different Administrative Levels	20
4.1.2 Determining Data Accuracy and Quality	21

4.1.3	Investigating Possible Sources and Methods for Collecting Unavailable Data	21
4.2	Implementation of Field Investigations	21
4.2.1	MALR Statistics Office, Governorate Level	22
4.2.2	Egyptian Survey Authority Office, Governorate Level	22
4.2.3	MALR Statistics Office, District Level	22
4.2.4	MALR Village Cooperative Office	23
4.3	Time Series Analysis	23
4.3.1	Introduction	23
4.3.2	Overview of Time Series Analysis	24
4.3.3	Time Series Analysis in This Study	24
5.	FINDINGS: COLLECTION PROCEDURES	26
5.1	Results of Team Visits	26
5.1.1	Village Cooperatives	26
5.1.2	MALR District Level Offices	27
5.1.3	Governorate Level Offices	27
5.1.4	General Observations	27
5.2	Data Available and Their Relationship To Data Needs	29
5.2.1	Existing Data Vs Available Data	29
5.2.2	Agricultural Data Collected by MALR	29
5.2.3	Data Requirements	30
6.	FINDINGS: TIME-SERIES ANALYSES	38
6.1	Introduction	38
6.2	Modeling Process with Illustrative Examples	41
6.3	Area, Yield, Production, and Cost of Production in Behira	47
6.3.1	Wheat	47
6.3.2	Rice	47
6.3.3	Maize	48
6.3.4	Cotton	48
6.3.5	The Time-Series Models of Total Cost of Production	49
6.4	Area, Yield, Production, and Cost of Production in Dakahlia	62
6.4.1	Wheat	62
6.4.2	Rice	62
6.4.3	Maize	62
6.4.4	Cotton	63
6.4.5	The Time Series Models of Total Cost of Production	63
6.5	Area, Yield, Production, and Cost of Production in Beni Suef	75
6.5.1	Wheat	75
6.5.2	Maize	75
6.5.3	Cotton	75
6.5.4	The Time Series Models of Total Cost of Production	76
6.6	Area, Yield, Production, and Cost of Production in Assiut Governorate	82
6.6.1	Wheat	82
6.6.2	Maize	82
6.6.3	Cotton	82

6.6.4	The Time Series Models of Total Cost of Production	83
6.6	General Results of Time Series Analyses for Data Quality and Accuracy ..	90
7.	FINDINGS: STATISTICAL ANALYSES	93
7.1	Graphical Investigation of Data	93
7.2	Estimates Moving from District to Governorate to Publication (D→G→P)	106
7.2.1	Number and Magnitude of Changes	106
7.2.2	Test of Paired Differences	106
7.2.3	Analysis of Variance	107
7.3	Estimates Moving from Cooperative (Village) to District to Governorate (C→D→G)	118
7.3.1	Number and Magnitude of Changes	118
7.3.2	Test of paired differences	118
7.3.3	Analysis of Variance	119
7.4	Cost of Production Data	131
7.4.1	Published Data of Total Cost of Production	131
7.4.2	Potential improvement of Costs of Production Data Quality	131
8.	CONCLUSIONS AND RECOMMENDATIONS	135
8.1	General Conclusions	135
8.1.1	Operation of Data Collection System	135
8.1.2	Data Quality, Availability and Utility	136
8.2	Recommendations	137
8.3	Suggested Improvements to Statistical Organizations	138
8.3.1	Restructuring Statistical Organizations	138
8.3.2	Statistical Training Program	139
8.3.3	Development of New Statistical Methods and Sampling Techniques	143
ANNEX	144

LIST OF TABLES

Table 1.1 : Effects of Major Reforms on Farmers, Government's Role and the Need for Statistics	2
Table 3.1: Sample Sizes of Commodities for Which Crop Cutting Surveys are Conducted	16
Table 5.1: Existing Agricultural Data	31
Table 5.2: Statistics of Use to Producers, Processors, Traders, Decision Makers, Policy Makers, and for Economic and Impact Analysts	35
Table 6.1: Code Numbers of Governorates and Districts	40
Table 6.2: Total Number of Time Series Cases by Governorate	40
Table 6.3: Area of Wheat in Behira Governorate	49
Table 6.4: Yield of Wheat in Behira Governorate	50
Table 6.5: Production of Wheat in Behira Governorate	51
Table 6.6: Area of Rice in Behira Governorate	52
Table 6.7: Yield of Rice in Behira Governorate	53
Table 6.8: Production of Rice in Behira Governorate	54
Table 6.9: Area of Maize in Behira Governorate	55
Table 6.10: Yield of Maize in Behira Governorate	56
Table 6.11: Production of Maize in Behira Governorate	57
Table 6.12: Area of Cotton in Behira Governorate	58
Table 6.13: Yield of Cotton in Behira Governorate	59
Table 6.14: Production of Cotton in Behira Governorate	60
Table 6.15: Cost of Production in Behira Governorate	61
Table 6.16: Area of Wheat in Dakahlia Governorate	64
Table 6.17: Yield of Wheat in Dakahlia Governorate	65
Table 6.18: Production of Wheat in Dakahlia Governorate	66
Table 6.19: Area of Rice in Dakahlia Governorate	67
Table 6.20: Yield of Rice in Dakahlia Governorate	68
Table 6.21: Production of Rice in Dakahlia Governorate	69
Table 6.22: Area of Maize in Dakahlia Governorate	70
Table 6.23: Yield of Maize in Dakahlia Governorate	71
Table 6.24: Production of Maize in Dakahlia Governorate	72
Table 6.25: Area of Cotton in Dakahlia Governorate	72
Table 6.26: Yield of Cotton in Dakahlia Governorate	73
Table 6.27: Production of Cotton in Dakahlia Governorate	73
Table 6.28: Cost of Production in Dakahlia Governorate	74
Table 6.29: Area of Wheat in Beni Suf Governorate	76
Table 6.30: Yield of Wheat in Beni Suf Governorate	77
Table 6.31: Production of Wheat in Beni Suf Governorate	77
Table 6.32: Area of Maize in Beni Suf Governorate	78
Table 6.33: Yield of Maize in Beni Suf Governorate	78
Table 6.34: Production of Maize in Beni Suf Governorate	79
Table 6.35: Area of Cotton in Beni Suf Governorate	79
Table 6.36: Yield of Cotton in Beni Suf Governorate	80
Table 6.37: Production of Cotton in Beni Suf Governorate	80

Table 6.38: Cost of Production in Beni Suef Governorate	81
Table 6.39: Area of Wheat in Assiut Governorate	83
Table 6.40: Yield of Wheat in Assiut Governorate	84
Table 6.41: Production of Wheat in Assiut Governorate	85
Table 6.42: Area of Maize in Assiut Governorate	86
Table 6.43: Yield of Maize in Assiut Governorate	86
Table 6.44: Production of Maize in Assiut Governorate	87
Table 6.45: Area of Cotton in Assiut Governorate	88
Table 6.46: Yield of Cotton in Assiut Governorate	88
Table 6.47: Production of Cotton in Assiut Governorate	89
Table 6.48: Cost of Production in Assiut Governorate	89
Table 7.2: Wheat Estimate Comparisons	109
Table 7.3: Rice Estimate Comparisons	110
Table 7.4: Maize Estimate Comparisons	111
Table 7.5: Fava Bean Estimate Comparisons	112
Table 7.6: Cotton Analysis of Variance (ANOVA) Results	113
Table 7.7: Wheat Analysis of Variance (ANOVA) Results	114
Table 7.8: Rice Analysis of Variance (ANOVA) Results	115
Table 7.9: Maize Analysis of Variance (ANOVA) Results	116
Table 7.10: Fava Beans Analysis of Variance (ANOVA) Results	117
Table 7.11: Cotton Estimate Comparisons Village Level	121
Table 7.12: Wheat Estimate Comparisons Village Level	122
Table 7.13: Rice Estimate Comparisons Village Level	123
Table 7.14: Maize Estimate Comparisons Village Level	124
Table 7.15: Fava Bean Estimate Comparisons Village Level	125
Table 7.16: Cotton Analysis of Variance (ANOVA) Results	126
Table 7.17: Wheat Analysis of Variance (ANOVA) Results	127
Table 7.18: Rice Analysis of Variance (ANOVA) Results	128
Table 7.19: Maize Analysis of Variance (ANOVA) Results	129
Table 7.20: Fava Bean Analysis of Variance (ANOVA) Results	130
Table 7.21: Variable Costs, Rent, and Total Cost of Production for Major Field Crops in the Selected Governorates as in 1997	132
Table 8.1: Summary of Training Courses and Number of Hours Needed	141
Table 8.2: Examples of Training Topics and the Appropriate Level of Trainees	142

LIST OF CHARTS

Chart 3.1: Organizational Structure MALR (October 1997)	11
Chart 3.2: The Organizational Structure of the Economic Affairs Sector	12
Chart 6.1: Illustration of Time Series Data Showing Nonstationarity in the Mean	39
Chart 6.2: Illustration of Time Series Data Showing Nonstationarity in the Variance	39
Chart 6.3: Illustration of Time Series Data Showing Nonstationarity both in the in the Mean and Variance	39
Chart 6.4: Original Series of Maize Yield in Assuit Governorate	43
Chart 6.5: Difference of Original Series of Maize Yield Data to Achieve Stationarity	44
Chart 6.6: The Autocorrelation Function of Maize Yield in Assuit Governorate During the Period 1970-1997	45
Chart 6.7: The Partial Autocorrelation Function of Maize Yield in Assuit Governorate During the Period 1970-1997	46
Chart 6.8: Sample ACF of most original time series under study	90
Chart 6.9: Sample PACF of most original time series under study	91
Chart 7.1: Cotton Area in Behira, Damenhour District-District, Governorate and Published	93
Chart 7.2: Cotton Production in Behira, Damenhour District-District, Governorate and Published	94
Chart 7.3: Cotton Area in Dakahlia, Mansoura District- District, Governorate and Published	94
Chart 7.4: Cotton Production in Dakahlia, Mansoura District-District, Governorate, and Published	95
Chart 7.5: Maize in Behira, Damanhour District-District Area/100, Yield, Production/1000	96
Chart 7.6: Maize in Behira, Damanhour District-Area/100, Yield, Production/1000	97
Chart 7.7: Maize in Dakahlia, Sherbin District-Area/100, Yield, Production/1000	98
Chart 7.8: Maize in Dakahlia, Sherbin-Area/100, Yield, Production/1000	99
Chart 7.9: Rice Area in Behira, Kom Hamada District-District, Governorate and Published Estimates	100
Chart 7.10: Rice Production in Behira, Kom Hamada District-District, Governorate and Published Estimates	101
Chart 7.11: Rice Area in Behira, Damanhour District-District, Governorate and Published Estimates	102
Chart 7.12: Rice Production in Behria, Damanhour District-District, Governorate and Published Estimates	103
Chart 7.13: Rice Area in Dakalahia, Mansoura District-District, Governorate and Published Estimates	104
Chart 7.14: Rice Production in Dakahlia, Mansoura District -District, Governorate and Published Estimates	105
Chart 7.15: Cost of Production -Wheat, Beni Suef Governorate	133
Chart 7.16: Cost of Production -Cotton, Beni Suef Governorate	133
Chart 7.17: Cost of Production -Maize, Beni Suef Governorate	134

LIST OF ACRONYMS

ACF	Auto-Correlation Functions
APRP	Agricultural Policy Reform Program
AR	Autot-Regressive
ARC	Agricultural Research Center
AERI	Agricultural Economic Research Institute
ARIMA	Auto-Regressive Integrated Moving Average
ARMA	Auto-Regressive Moving Average
ATUT	Agricultural Technology Utilization and Transfer (USAID funded project)
CAAE	Central Administration for Agricultural Economics
CACU	Central Agricultural Cooperatives Union
CAAP	Central Administration for Agricultural Planning
CAPMAS	Central Agency for Public Mobilization and Statistics
CBE	Central Bank of Egypt
DF	Degrees of Freedom
EAS	Economic Affairs Sector
ESA	Egyptian Survey Authority
F1	First Difference
F2	Second Difference
FAO	Food and Agriculture Organization (UN)
GARPAD	General Administration for Reclamation, Projects, and Agricultural Development
GIS	Geographic Information System
GOE	Government of Egypt
GTZ	Deutsche Gesellschaft fuer Technische Zusammenarbeit
IAP	Impact Assessment Plan
IFPRI	International Food Policy Research Institute
LN1	(Log) _e
MA	Moving Average
MALR	Ministry of Agriculture and Land Reclamation
MPWWR	Ministry of Public Works and Water Resources
MS	Mean Squared
MTS	Ministry of Trade and Supply
MVE	Monitoring, Verification and Evaluation Unit (APRP)
MW	Yield per Feddan
PACF	Partial Auto-Correlation Functions
PBDAC	Principal Bank for Development and Agricultural Credit
RDI	Reform Design and Implementation (APRP Unit)
SS	Sum of Squares
TS	Time Series
USAID	United States Agency for International Development
USDA	US Department of Agriculture

ACKNOWLEDGMENTS

The authors would like to acknowledge the wise guidance and scientific advice of Dr. Saad Nassar. Appreciation is also extended to Eng. Mahmoud Nour for his strong support during the study implementation. Special thanks go to Drs. Mohamed Omran and Glenn Rogers for the long discussions, feedback, and their participation in designing and implementing this study. The authors feel really in debt to Dr. Gary Ender the COP of MVE for his guidance, participation and support in formulating and implementing this study, and reviewing the report.

Appreciation should be extended to Eng. Mohamed El Shahed, the head of the Central Administration of Agricultural Economics, and Eng. Abd El Razik Hassan, the head of the Central Administration for Agricultural Planning. Without their cooperation and assistance, this study would never have been completed.

The authors also would like to acknowledge the hard work and dedication of the researchers who conducted the field investigations: Dr. Mohamed Messelhy in Dakahlia, Dr. Mohamed Heikel in Behira, Dr. Mohamed Saied in Beni Suef, and Dr. Mahmoud Alaa Abd El Aziz in Assuit.

Special thanks go to Mrs. Yvonne Louis, Miss Flora Naiem, and Mr. Hisham Amin, for their hard work in data entry, tabulation, and file manipulation.

Finally, any errors and omissions should be attributed to the authors and not to the APRP, USAID, or MALR.

EXECUTIVE SUMMARY

The structure of farming in Egypt has totally changed over the past 50 years. It has gone from a small number of very large holdings, managed by land owners, with little government control; through break-up of holdings and total government control; to a very large number of small holdings, independently managed, with the Government taking on an advising and assisting role. Along with these structural changes the need for information and statistics has changed radically: from little need for information; to accounting type data to assist a few decision makers; to very detailed and complex data on all aspects of the agricultural economy to assist millions of decision makers.

The **main objectives** of this study are to:

- Review the Impact Assessment Plan (IAP) and compile a list of the data needed for the analyses in the plan,
- Identify the available agricultural data at different administrative levels,
- Determine the accuracy and quality of the available agricultural data, and
- Investigate possible sources and methods for collecting unavailable data.

The **methods** adopted in this study are:

- Assessing the performance of the original sources of agricultural data,
- Diagnosing the procedures from initial data collection to publication, and
- Assessing the quality of the published time series agricultural data (MALR's published data) by applying time series techniques.

The study team went to 4 governorates, 2 in the Delta (Behira and Dakahlia), one in Middle Egypt (Beni Suef) and one in Upper Egypt (Assuit). In each governorate, three districts were visited, and two villages within each were also visited.

The current study examined the data collection process in more detail than past studies, and accomplished the following unique work:

- Following the data from the source villages, through district and governorate offices to the national office; determining if changes are made, where, and why.
- Studying in detail the data gathering process, identifying problems and constraints to good data handling at each level.
- Creating an electronic database at the district level for four governorates.
- Conducting time series analysis on the MALR data series to see what the data can tell about the process.
- Testing possible methods for collecting data that are currently not available.
- Looking in detail at the data available to see whether they are accessible and reliable.

Findings: Statistical Organizations

Sampling Offices in the Governorates. The problems common to all governorate sampling offices were found to be:

- Old equipment that is heavy to carry and hard to use and causes some errors.
- Some villages are hard to reach because they have few cars and old motorcycles.
- Communications among office, village and farmer are difficult.
- Hard to determine right time to visit farmer.
- Old sampling frame maps and listings.
- Lack of proper training.
- Many staff nearing retirement age; no new staff learning the work.
- Office space very cramped.
- Low pay.

Statistical Offices in the Governorates. In these offices we found:

- Lack of cooperation between Agricultural Affairs staff and Sampling staff.
- A reward is given to Agrarian Affairs staff in the governorates based on their crop yields. This puts a lot of pressure on leaders to influence their yield figures.

Statistical Organizations. In general we find:

- No clear line of authority or responsibility.
- Lack of technical direction.
- Lack of support: facilities, supplies, filing systems, transportation, per diem
- No standardized forms or procedures.
- The Egyptian Survey Authority does much work and is somewhat independent of the Agricultural Affairs work, but there is much duplication of effort between these two agencies. Significant savings could probably be realized if there were a restructuring and improvement of the area estimation process.

Findings: Time Series Analyses

Area, Yield and Production Data. Most of the 182 sets of data analyzed were for the major field crops by district and governorate. One would expect to find different data patterns for different crops in different administrative areas. It was surprising to find most of the data sets with:

- The same type of nonstationarity in their time series data,
- The same time series model AR (1),
- Very low levels of forecast errors.

All crop data were too uniform, indicating some controlling influence, and thus leading one to doubt the accuracy of the data. However, cotton data did appear to be more accurate than the others.

Cost of Production Data. Total costs of production are obviously affected by: The prices and quantities of the inputs used in the production process; The level of output produced; and The state of technical knowledge used in production. The directions and variations in the above three factors are not the same. Therefore, it was very unexpected to find that the time series data of total cost lay on a straight line. This unusual result held true for total cost data of wheat, rice, maize, and cotton. Moreover, the total cost time series for all four crops had the same surprising results as the area, yield and production data discussed above. Therefore, we can say without hesitation that the total cost time series data are too uniform and thus likely not accurate. Again, some interference is indicated during the collection and reporting processes of these data.

Findings: Village, District and Governorate Data

Estimates Moving from District to Governorate to Publication. Most statistical systems would expect less than 1% of estimates to be changed, and this only with detailed explanation. The percent of estimates that were changed, as they went from district to governorate or publication, ranged from 36% to 96%, with most being above 60%. This is an extremely high number of changes. The number of changes appear to be more moderate for major field crops like cotton and larger for minor crops like maize. This is probably due to less check data and more diverse use of these crops.

Statistical *paired difference tests* were made to determine if the changes in data were unusual. For most of the results the null hypotheses of no changes in the estimates as they go to higher levels is rejected. A non-parametric *Wilcoxon Signed Rank Test*, which looks at the number of positive changes and negative changes and tests these two numbers, was also performed. The results of these tests were essentially the same as the paired difference tests. As data goes from lower to higher level it takes on higher values. *Analysis of Variance* tests were made on the data to determine if data varied within and between levels, and if so, where. Almost all of the tests had F-values so great as to have a probability of less than 0.0001 of a greater F-value.

Estimates Moving from Cooperative (Village) to District to Governorate. Generally the C→D→G data were more sparse than the D→G→P data, but were still adequate for the analysis procedures. The percent of estimates that were changed was about the same as in the previous section. The magnitudes of change were less, as one would expect. Interestingly, there was only one data set found at the village level, which is in sharp contrast to what we found at the district level. Over half of the paired difference tests showed some level of significant data changes. Most of the analysis of variance tests were significant.

Findings: Data Availability

Different data are available in governorates and districts. This indicates that slightly different procedures are used in these locations. For example:

- The most valuable source, the village extension agent, is not being adequately acknowledged or utilized, which decreases data availability.
- In Behira, village-level data were only available for the latest 1 or 2 years. District

offices had village data for most years. Governorate offices had village data for the latest 4 years.

- In Beni Suef, district and governorate offices had fairly complete village-level data, probably due to carbon copies of village data being sent to these offices. If this is the case, there were still some changes being made at the higher levels.
- In Assiut, village-level data available at village and district offices are fairly consistent. Governorate offices do not have village data.
- In Dakahlia, for some villages, data were too uniform for area, yield and production. This makes one suspect that constants or technical coefficients have been used instead of current estimates. If so, the estimation may have actually taken place at the district or higher level without consulting with village agents.

Findings: Data Quality

For the major field crops, the team finds that:

- Area data are fairly reliable if obtained at the village level, before changes are made.
- Yield estimates of the sampling offices are good.
- If these crop yield estimates are applied to valid area estimates, then reasonably good production estimates may be derived.
- All other data, including cost and return data, are suspect. Cotton data appear to be of better quality, due to the emphasis placed on them.

Agricultural Sector Environment and its Effect on Statistical Information and Quality.

The statistical, sampling and area measurement office staffs are to be commended for trying to carry out their mandates under often very difficult conditions with little or no administrative support. At all levels, the staff are doing work for which they neither had training nor were given any support or recognition. The statistical work is considered of minor importance by the village extension staff, especially since they receive no incentive payment for it. Most offices have plans for quality control procedures, but admit that these often have to be bypassed due to equipment or budgetary constraints. The manipulation of data as it travels up through higher levels is very prevalent and thus accuracy is doubtful. There seems to be a governmental disincentive to accurate information. Many staff tell of their data being changed so that higher level government officials could receive awards or recognition for high yield or production.

Data Quality and Usefulness for Impact, Economic and Policy Analysis. The team finds that:

- The current statistical system is inadequate to meet current and future information and data analysis requirements.
- The time lag between data gathering and publication is too great to meet modern requirements. Proper data handling techniques and use of computers would make information more timely.
- In spite of the conditions in the village cooperatives, the extension agent records and maps do contain excellent *area* estimates. These extension agents live and work in the villages on a daily basis. Some of them are farmers themselves. They have created

maps and have farmer areas listed on sheets or in books. Cotton receives special emphasis.

- Some of the data available at the village is not forwarded upward through administrative levels, and other data are forwarded only as requested. The village cooperative staff have very detailed information that could be used for sampling frames for economic and research purposes.
- The best *yield* data come from the sampling offices, which conduct crop cutting surveys at harvest time.
- Many data needed for decisions in a market economy are missing. A comparison of the items available with the items needed makes the data gaps very obvious. A few items are available at the governorate or national level, but are needed at much lower levels, i.e. per feddan, per farm, by farm type or geographical region.
- There are estimates of *cost of production, farm gate prices, labor and wage rates* available for governorates. However, our investigation found that most of these were very subjective and covered a very narrow segment of the sector. Many of the village extension agents had such data, but stated that they were not asked for it. The governorate level data are often gathered at the ministry office from staff who are also farmers. Their costs are hardly those of the farmer on less than one feddan.
- *Farm gate prices* (from three farmers in each village) were found to be very different from the published governorate prices. They also were different from village to village.
- The implication for organizations doing impact, economic and policy analyses is that they will have to design and execute their own surveys, at least until a national statistical program can be established.
- Accessing specific information is often very difficult. Some basic and intermediate data are only kept for a year or two due to storage or records maintenance problems. Most data is on handwritten data forms, and data management is cumbersome.

Recommendations

While the data quality may not be very good at present, the prospects for the future could be bright. There is an existing infrastructure that could be built upon to develop a much improved statistical organization. The extension agents, with a little support and training, could be a valuable source of current and reliable data. The district, governorate and national staff, with proper training and support, could be safe conduits for the data to be aggregated and disseminated.

Restructuring the Statistical Offices in the Governorates

The current and future needs for information are increasing at an exponential rate while the current capability to provide the information is nearly static. Government must act quickly to create an environment for accurate, timely statistics to develop. Each decision made by farmer, trader, importer/exporter or government policy maker needs good information on which to base his decision. Presently little data is available, so each has to gather his own or make his decision in an information vacuum. There is an urgent need for assistance to the Government of Egypt to facilitate the creation of an environment and structure for high quality statistics. Significant savings could probably be realized if there were a restructuring

and improvement of the area estimation process.

There is a need for delegation of specific responsibility and authority to each office and level of government. This may be done by issuing a decree to assure protection of the statistical organization from external influence and the confidentiality of farmer information from exposure to taxing or regulatory agencies.

Statistical Training Program. Technical training to staff will be required to make the transition effective. Some support with equipment and supplies will also be advantageous. Based on perceived needs, the team recommends a program that provides for comprehensive training; the details are in section eight. For general training, all five courses should be completed in sequence. This affords coverage of all important aspects of statistical work. However, each course can be offered independently to suit specific training needs. Also, course organization allows for modifications that may be found necessary to suit the specific personnel to be trained, such as condensing or expanding specific topics. For a particular investigation, such as an agricultural census or sample survey, these courses would be suitably expanded and supplemented to provide for their special skill. This program emphasizes manual processing rather than electronic data processing, since there are still a number of manual processing operations involved before data processing can be passed on to the electronic data processing facilities.

Developing New Statistical Methods and Sampling Techniques. Most of the procedures followed in collecting data need to be improved. The EAS has already started this process by constructing an experienced team. They need to redesign questionnaires. In addition they need to redesign the intermediate tables, pre-formatted tables, and the final tables to be sent to the higher levels. Of course, statisticians need to be trained on how to use these materials.

Extension agents have reasonably good estimates on production and yield, as they talk to farmers frequently and can make observations. During discussions with farmers about cultural or pesticide problems, they can get information about farmer costs and returns. Farmers sometimes do not speak freely to strangers, but will confide in the extension agent, who is often a member of his village.

Clearly a better system of cost and price gathering is necessary. A village level survey could be instituted to give much more representative and reliable information. With some training and support, extension agents could provide good cost of production, farm gate prices, estimate volumes produced, and many other types of data.

The Sampling General Directorate needs to update the current sampling frame they use to select samples. The sampling techniques used currently need to be reviewed. The new generation of statisticians who work in sampling activity need to understand the theory of sampling and its applications in the field.

In adopting these new techniques, the following steps should be followed:

- Review procedures and survey methods used elsewhere in the world;
Determine if applicable to use in Egypt,

How could they be adapted for use in Egypt,
Do research to develop and test methods.

- Review literature to determine if new research holds potential methods that could be beneficial to Egypt's statistical program.
- Set up research to develop parameters for forecasting models. It would be advantageous to start the objective yield work again to give forecasts of yields early in the season.

1. INTRODUCTION

Egyptians have always been record keepers. Ancient hieroglyphic writings tell of agricultural production and harvests during the times of the pharaohs.

1.1 Agricultural Policy and Statistics

Table 1.1 highlights changes in the agricultural sector over the last 50 years. It also highlights the affects of agricultural policy on the farmer activities, government functions and need for statistics.

As one can see, the structure of farming has totally changed over the past 50 years. It has gone from a small number of very large holdings, managed by land owners, with little government control; through break-up of holdings and total government control; to very large number of small holdings, independently managed, with the Government taking on an advising and assisting role.

Along with these structural changes, the need for information and statistics has changed radically: from little need for information; to accounting type data to assist a few decision makers; to very detailed and complex data on all aspects of the agricultural economy to assist millions of decision makers.

To meet the current and future needs for information, one can build upon the infrastructure that was established during the governmental control phase. During this time each village was assigned extension agents to tell farmers what to grow, help them get their needs from the cooperatives or PBDAC (such as seeds, fertilizer, pesticides), and to channel their production to Government processors and exporters. Farmers had land but were essentially working for the Government. Extension agents were administrators and enforcers of government policy, and as such, accounted for the production of mandated crops and livestock.

Under the privatized system, farmers have a choice whether to follow government suggestions or not. To make decisions, farmers need to know the expected costs of inputs, the local market prospects, and any international effects on these markets to determine the profitability of these crops. Government policy makers need to understand the decision processes of farmers, suppliers, traders and the economic environment. Extension agents now need to function as educators and facilitators of new technology. Extension agents can still gather area, yield, production, costs and price data by observation, measurement and discussion with sector enterprises. However, much critical data for policy and decision makers can only be obtained from the farmer, i.e. family labor costs, actual costs for inputs, farm gate prices received, and other factors affecting his decision on what to produce and when.

Table 1.1 : Effects of Major Reforms on Farmers, Government's Role and the Need for Statistics

Timeframe	Farmer	Government	Statistics
Prior to 1952	95% of land owned by 5% of population. Largest 1% of population workers on farm,	Little control over farming	Large farmers had information system. Little public statistics
Sept. 1952 1 st Agricultural Reform	Agricultural land ownership 200 feddans per owner (300 per family). Workers on farms given 5 feddans (8 per family)	Control of farming decisions. Assigned crop type give inputs; require production quota for main crops. All marketing of output	records of transactions, i.e. areas assigned to farmers, inputs given, amounts received from farmers.
1961 2 nd Agricultural Reform	Reform reduced to 100 feddans per owner, up to 200 feddans per family. Workers on farms given 5 feddans.	Complete delivery of cotton and sugarcane production to government plants.	over all data from land reform areas. Mostly used for government planners and administrators
1969 Later Reform	Ceiling of 50 feddans per owner and 50 for the family	“	“
.	.	.	.
1986 privatization process begins	Farmer has to start making decisions about his operation again.	Reduction of subsidy on fertilizers, seeds and pesticides	Basic data on affects of how farmers and economy are reacting to privatization.
1997	Farmer making most decisions .	Subsidy for pest control for cotton. Sugarcane through prices paid.	Data requirements increased.
Currently	Since the 50's the land has been divided within the generations until 60% of the holders have less than 1 feddan, another 20% of holders have between 1 and 3 feddans. Farmers choose what to produce, buy inputs (seed, fertilizer, pesticides, equipment), sell produce	Government recommends crop areas, but does not command. Government assists by providing information.	Data requirements are many and critical. Area, yield, production, farm gate prices, costs at all stages of production, marketing, processing, import and export.

1.2 Organizational Structure of Agricultural Data Collection and Publications

Within the Ministry of Agriculture and Land Reclamation (MALR), the Economic Affairs Sector (EAS) is the main department responsible for collecting, tabulating, and publishing agricultural data. (Refer to Charts 3.1 and 3.2 for the organizational structures.) This department makes data and statistics available for all types of users. The EAS consists of two main divisions: the Central Administration for Agricultural Economics (CAAE) and the Central Administration for Agricultural Planning (CAAP).

The CAAE collects, tabulates and publishes current statistics. It contains three main general directorates: Agricultural Census, Agricultural Statistics, Food Security Projects, The CAAP contains the four main general directorate :Agricultural Finance, Agricultural Economic Resources, Agricultural Policy, and Planning and Monitoring Agricultural Investment Projects. Between them these departments produce statistics on agricultural production, farmgate prices, costs of production, and other similar data. The CAAP contains most of the available computers for the EAS. Its main function is to process data and maintain the required databases.

For current agricultural statistics, each statistical department at the governorate level produces its own statistics from reports of agricultural extension agents. The MVE Impact Assessment Planning (IAP)Team conducted a field trip to examine how the collection system for agricultural statistics operates at these levels. Based on visits to two districts in two separate governorates, the descriptions were not consistent between the district and the governorate levels, so these impressions were taken as a point of departure for the current investigation.

In addition to the current statistics, the agricultural sector collects data via the agricultural census, annual crop cutting surveys, regional reports, estimates obtained from persons with professional experience. The EAS also tabulates and publishes data produced by other departments of MALR and by other ministries (MTS, MPWWR, PBDAC, CBE) and organizations (CAPMAS). Such data include, for instance, estimates of cultivated area of fruits and vegetables, data on livestock and international trade for agricultural commodities. One of the main publications of the EAS is the “Annual Bulletin of Agricultural Economics”, the main source of official agricultural statistics in Egypt.

The potential usefulness of the current system, if its accuracy can be confirmed, coupled with the relatively low cost of modestly expanding the data collected by the Egyptian Survey Authority, certainly warrants a more comprehensive investigation into the MALR data collection system to assess the quality of its data on a wider scale.

1.3 Objectives of the Study

In response to the new and detailed data requirements for decision makers from government to farmer level, one needs to know what information is available, the quality of data and how it is obtained. This work was undertaken to assess data availability and quality.

The main objectives of this study are:

- Review the MVE Unit's Impact Assessment Plan (IAP) and compile a list of needed data for various proposed activities of the impact assessment plan.
- Identify the available agricultural data at different administrative levels.
- Determine the accuracy and quality of the available agricultural data.
- Investigate possible sources and methods for collecting unavailable data.
- Suggest possible sources and methods of improving data quality.

1.4 Outline of the Study

The remainder of this report is divided into seven additional chapters. Chapter two describes the criteria used in assessing the quality and availability of agricultural data. This chapter also includes the types of errors made in collecting data. Chapter three covers the main sources of data collection in the agricultural sector and their organizational structure. This chapter also includes a description of the data collection and estimation procedures.

Chapter four discusses the methodology followed in this study. This chapter includes the approach followed in investigating the statistical offices at all levels and the types of models applied in the time series analysis of the published data.

Chapter five covers the main findings regarding data collection procedures. It includes a summary of the investigations conducted during the team's field trips. This chapter also contains a comparison of the available data with these required.

Chapter six reports the results of the time series (Box- Jenkins) analysis of the published data of four governorates, with a summary of the main findings.

Chapter seven covers the main findings of the statistical analysis. It contains all of the results of the comparison between the published estimates and the estimates collected through the investigation of the statistical offices at different levels.

Chapter eight contains all of the conclusions and recommendations of this study. It contains some specific suggestions in the area of training, restructuring of the statistical offices, and proposed procedures to be followed in the future.

2. CRITERIA FOR DATA QUALITY AND AVAILABILITY

Before one can determine data quality it is necessary to define some terms which may be used as criteria against which to judge. The following discussion will provide some suggestions. It is not intended to be a complete discussion, but will cover the main characteristics.

2.1 Types of Data

2.1.1 Time Dimension and Intended use

The timelines required of data to be useful depends on the use to which they will be put.

Historic data is very important to explain past circumstances and conditions that existed and why certain things happened. Sociologists and historians can explain many things by studying past information. Censuses and other historic data series perform an invaluable service when they are accurate. Where the current situation is similar to past conditions, historic data can be used to predict what may happen during the current season. Historic data is of little value when the current situation is not like the past or the past data cannot be reasonably modified to simulate the current situation.

Historic data can serve as benchmarks for judging progress or success over time. Censuses perform valuable service as benchmarks, and often give very detailed information on slowly changing characteristics of the population. However, the time required to process the volume of data prevents censuses from providing information on rapidly changing conditions or results that are needed quickly.

Current data relates to information that describes the current situation. Farmers need current market prices to help them decide in which market to sell their produce and the best time to sell. Government leaders need to know what the current production is to decide whether to buy grain or sell grain on the international market. Current data is critical for the day to day decision process at all levels of the economy.

A third type of data tries to look into the **future**. A **forecast** tries to tell what is going to happen in the near future, i.e., the production of cotton at harvest time, three or four months before harvest actually begins. A **prediction** tries to tell what is going to happen in the distant future, i.e., what will be the market for cotton in five years. Forecasts are very helpful for emergency warnings of impending drought or needed purchases on the world market for shortages in the coming crop year. Predictions are important for planning policies that will bring the production to a level desired in the future markets. Forecasts and predictions rely on relationships between historic data and current field and plant measurements. It usually takes some time and research to establish the valid cause and effect relationships that can be measured to provide forecasts and predictions.

2.1.2 Quantitative vs. Qualitative Data

Qualitative data denotes a characteristic or specifies a level or a condition.

Traits like gender, ethnic group, animal breed and plant specie are examples of characteristics variables. Treatment levels in experiments like fertilizer, temperature and pesticide applications are recorded as qualitative variables. Conditions like sunny, rainy, presence or absence of an economic or social factor are coded as qualitative variables. Qualitative variables can be designated by either alphabetic characters, symbols or discrete numbers.

Quantitative data usually relates to measurements or functions of continuous data.

Examples are field measurements, production, feddans of cotton on farm, ardabs of wheat per feddan. Quantitative variables are usually designated by continuous numbers, with decimals to the desired precision.

The summarization and analysis procedures, and statistical tests can be very different depending on whether the variable is quantitative or qualitative. Therefore, it is important to identify the type of data so that appropriate summarization and analysis can be done. With current computer capability it is possible to do inappropriate analysis and come to incorrect conclusions if type of data is not identified properly.

2.2 Types of Errors

It would be ideal if all of the data did not have any errors. If data were exactly correct for each farm and all farms were tabulated without mistakes, then estimates for all variables would be the true values. Unfortunately, this utopian condition seldom exists.

2.2.1 Non-Sampling Errors

occur due to causes other than sampling. If one makes an error in measurement, recording, copying, or tabulating, he has made a non-sampling error. These are often mistakes due to accidental equipment or human failure. There are also intentional errors introduced by buyers or sellers lying about their costs and prices, or politicians modifying data on area or production in their domain of interest.

A certain amount of non-sampling error will occur under the best of conditions, but frequently occur in large, loosely controlled data gathering and processing operations. These errors are hard to detect and control. Quality control measures can be used to determine and control errors, but are often difficult to implement and often the first to go in budget cutting measures. Conceivably, non-sampling errors could be so high in large surveys, without proper quality control, that the survey results are invalid.

2.2.2 Sampling Errors

occur when sampling procedures are used instead of complete enumeration. These errors are inherent in the process, but their magnitude and direction can be determined and controlled by the process. Thus when sampling procedures are used to make estimates, a statement can be made as to how precise the estimate is and one can then determine the reliability he is

willing to place on the estimate.

Care should be taken when designing a survey to assure that the sampling procedure is appropriate and will provide reliable estimates. The potential sampling frame often determines the most effective sampling procedure. The advantage of the sampling process is that only a small representative part of the population is used for data gathering, and strict controls can be followed to eliminate virtually all non-sampling errors. The smaller amount of data can be processed and published in a short amount of time and the precision can be stated.

2.2 Non-Response Errors

The non-response errors are due to missing data or misrepresentation of a part of the population. If any person in the survey does not respond, his data is missing and the estimate will be in error unless proper adjustments are made. Especially for voluntary response surveys, whole areas or ethnic groups might not respond due to superstition or for cultural reasons. Their characteristics and data levels could be different from the other members of the population. It is important to check for non-response errors. For example, great efforts are made to be sure data is gathered for the large farms, and one may forget to be sure that the smaller farms are properly represented. If the cost of production information only represents the large farm costs, then the preponderance of struggling small farmers are ignored and policy decisions based on survey results will be flawed.

2.3 Criteria For Evaluating Information (Estimates or Forecasts)

The criteria below can be used to evaluate the quality of data. However, the type of data and potential errors discussed above also affect the determination of data quality. The criteria will first be described and then some comments about the interaction between factors will be made.

2.3.1 Timeliness

Can be defined as having the information available when it is needed. Information in November about this year's cotton production is adequate for a historian and may be in time for a buyer for cotton gins, but is too late for the importer who needed to know there was going to be a shortage or surplus four months before harvest began. Thus the data was timely for the historian, and maybe the gin buyer, but was worthless for the importer. A good statistical program meets all user's requirements in a timely manner.

2.3.2 Completeness of Coverage

For the population of interest is an important consideration. One would like to assure that all parts of the population are represented properly in statistical information. It is understandable that large farming operations are included in the data and resultant statistics. Often a large number of small farmers are ignored or not properly represented in the statistics even if they may produce a substantial part of the production. These data gaps sometimes occur due to time, cost or difficulty considerations, but result in incomplete coverage of an important part

of the agricultural sector. It is not unusual for statistical organizations to exclude a small part of the population from their statistics, but this is only valid if they have first proven that they make up only a miniscule part of the population. For example, the USDA excludes small operations less than a specified acreage, but they know that their production amounts to less than 1% of all commercial production.

In Egypt there are a large number of small farms, which make up an important part of the agricultural economy. Their costs, returns and economic condition are very important to agricultural policy and impact considerations. Data on their operations are critical and should not be ignored or underrepresented in statistics. Thus a system should be devised to make sure that these are given proper representation in the coverage. This can be done through proper sampling.

2.3.3 Accuracy

The true value of the population (the exact wheat area planted at some point in time or exact number of cattle on a given date) is unknown. If we knew it, we would not be trying to make an estimate. If the estimate of a population value is exactly the true value then the estimate is *unbiased* and *accurate*. This is seldom the case, since even censuses have errors and do not equal the exact population value.

Bias is defined as how far the estimator is from the true value. If our estimator is biased but will give us a value close to the true value, it is still very important to us. If the bias is small, then knowing the estimate is almost as good as knowing the true value. Often the true value is impossible or very expensive to obtain, while an estimate is much easier and cheaper to obtain.

Precision is a measure of how close the estimation procedure is expected to be to the true population value. It is usually expressed in terms of a range about, or percent of, the true value. With known estimation procedures it is possible to compute the error and bias of the process and thus estimate the true value precisely.

2.4 Evaluating Data Quality

Includes consideration of the above data types, potential data errors and criteria in light of data needs. If the information provided does not meet the **data users requirements** it is of little value to him no matter how accurate, timely or complete it may be. Data users have specific data requirements that may include different levels of aggregation and different frequencies. For example, government planners may be satisfied with governorate information on a monthly basis; while a trader or supplier may need daily or weekly information at a district or village level; a researchers may need farmer or field level data at the time of each cultivation and harvest operation. Data must be easily available in the format that the user can use. Traders and suppliers may want printed information while the researcher or economist may want data on diskette.

Comments about interaction between factors:

- The timeliness required of data depends on the use to which they will be put.

- Completeness of coverage might be affected by non-response or non-sampling errors.
- Accuracy could be affected by errors in survey procedures.
- Inappropriate analysis due to miss-identification of variable type will lead to wrong results.
- Large non-sampling errors can invalidate survey results

2.5 Methods to Improve Data Quality

2.5.1 Quality Control

Procedures can be instituted to reduce or eliminate non-sampling and non-response errors. Questionnaires can be pre-tested to make sure that the correct information is being obtained. Instructions and check data boxes can be included on the questionnaires. Supervisors can check all of the enumerator's questionnaires and randomly check their fieldwork. Re-interview research and objective field checks can identify if respondents are giving incorrect data. During data processing any variable value that exceeds certain limits can be verified before being processed and consistency checks can be made within respondent's questionnaire to reduce errors. To be of value the quality control measures must be designed to improve quality and not be a mere repeat of the original operations.

2.5.2 Check Data

Sometimes one is fortunate and has a source of final area, production, export, or market information which can be used to "check" or "true-up" early season estimates. (Cotton ginning outputs have been such a source in the past). When check data is available then it can be used to improve estimation formulas or relationships. Through the years more precision can be built into the process. Unfortunately, Egypt does not have many such sources. Most of the production is used on the farm or it goes through informal channels or is processed in factories which do not accurately report their outputs.

In such cases where true production levels are not known, then a properly designed and executed survey may be the only way to get accurate and representative statistics about the population. Sampling and survey procedures have been developed, used and proven to be effective in many countries around the world. Sample surveys have been found to be more cost effective and precise than total coverage, which is prone to many errors.

3. COLLECTION OF AGRICULTURAL DATA IN EGYPT

3.1 Structure of Agricultural Data Collection Organizations

Within the Ministry of Agriculture and Land Reclamation (MALR), the Economic Affairs Sector (EAS) is the main department responsible for collecting, tabulating, and publishing agricultural data. (Refer to Charts 3.1 and 3.2 for the organizational structures.) The EAS consists of two main divisions: the Central Administration for Agricultural Economics (CAAE) and the Central Administration for Agricultural Planning (CAAP). It contains seven main general directorates: Agricultural Census, Agricultural Statistics, Food Security, Agricultural Finance, Agricultural Economic Resources, Agricultural Policy, and Planning and Monitoring Agricultural Investment Projects. Between them these departments produce statistics on agricultural production, farmgate prices, costs of production, and other similar data. They make statistics available for all users.

Within the CAAE, the general directorate for Agricultural Statistics is responsible for collection, tabulation and publication of all current statistics. This directorate has statistical departments at each governorate and statisticians at every district which produce their own statistics from reports of agricultural extension agents in villages and cooperatives. The accumulation and tabulation of these data is the basis of their statistics. Also within each governorate there is a sampling office which does the crop-cutting surveys and other research requiring measurements in the field. These offices have just combined administratively, but their functions have not yet been integrated. The CAAP contains most of the available computers for the EAS. Their main function is to process data and maintain the required databases.

In addition to the current statistics, the EAS collects data via the agricultural census, regional reports, expert judgments from persons with professional experience, and some computed indications. The EAS also tabulates and publishes data produced by other departments of MALR and by other ministries (MTS, MPWWR) and organizations (PBDAC, CBE, CAPMAS). Such data include, estimates of cultivated area of fruits and vegetables, data on livestock and international trade for agricultural commodities. One of the main publications of the EAS is the “Annual Bulletin of Agricultural Economics”, the main source of official agricultural statistics in Egypt.

Another source of area data is through the Ministry of Public Works and Water Resources (MPWWR), Egyptian Survey Authority (ESA). They have independent offices at governorate and district levels. One of their major functions is to measure a sample of cotton and wheat fields to verify area planted to these crops.

Chart 3.1: Organizational Structure MALR (October 1997)

Chart 3.2: The Organizational Structure of the Economic Affairs Sector

3.2 Description of Data Collection and Estimation Procedures

The previous section shows the organizational structure of the MALR and EAS. It mentions some of their functions. This section will discuss the main statistical organizations, their activities and inter-relationships, if any.

In any organization there are the theoretical or stated procedures, and the actual procedures which are followed. We will attempt to describe the theoretical operations of each organization, the actual way things are done, and the problems encountered that often cause the differences between the two procedures.

3.2.1 Agricultural Census, EAS/MALR

Information obtained from the agricultural census is the most detailed data available on the Egyptian agricultural sector. There are many types of information for which the census is the only source. Many census items are of interest, but of minor importance relative to commercial and export needs. The census also provides benchmark information for types and magnitudes of production and details on crop, livestock and housing.

Egypt started conducting an agricultural census immediately following the international agreement in 1928. Since that time, six agricultural censuses have been conducted: in 1929, 1939, 1950, 1961, 1982, and 1990. All of these were carried out two years after conducting the population census. This provided an up-to-date sampling frame for the agricultural census, at a considerable savings in time and money.

As with most censuses, timely summarization and publication were a real problem. It took 6 to 9 years to summarize and publish the complete agricultural census. This time lag made the publication mostly beneficial for historic purposes. However, the EAS starts currently to reduce this period in the pilot census of year 2000 agricultural census.

A preliminary summary of key data items at the village level is made by the census department. Some research was done with these summaries to determine if they might be a resource for stratification of villages. This could be a potential source of stratification in development of a sampling frame for statistical surveys.

Dissemination of census results and publications is very limited due to publication cost and timeliness considerations.

3.2.2 Current Agricultural Statistics, EAS/MALR

The Current Agricultural Statistics office is responsible for accumulating all information and estimates on the agricultural sector and publishing them. It publishes data on crops, livestock, costs and returns, imports and exports. Section 5.2.2 discusses the data available and Table 5.1 gives a detailed list.

Current agricultural statistics are gathered through the governorate and district Agricultural Affairs Offices. The agricultural districts are covered by extension agents, each of whom has

150-300 feddans to follow. The agents advise farmers and collect data about the major field crops (input and output data). At the appropriate reporting period the agents summarize information on their farmers and pass it to the district level. Their coverage is supposedly a census of all farms producing a specific crop. Theoretically, the extension agents gather all the data from farmers, which are then accumulated and passed through the district and governorate offices to the national headquarters without manipulation or changes. Thus the information is felt to be complete, correct and true.

3.2.3 Egyptian Survey Authority, MPWWR

The Egyptian Survey Authority (ESA) is responsible for area statistics for all economic sectors in Egypt. Of special interest in the agricultural sector are their measurements of fields. The ESA is responsible for measuring areas of cotton and wheat within sample hodes. They report these to the EAS in MALR who compares these with their estimated areas. When large differences occur, the sampling office takes a sample to determine which estimate is more correct.

The team visited a governorate office where we were given the theoretical procedures. They measure 50% of the cotton and wheat area in the districts. They send out a team of four people (technician, chief, inspector and assistant Inspector). The technician draws a map of the hode and measures all fields. The chief rechecks the technician's drawing and 25% of the field measurements. The inspector and assistant inspector do a 25% re-measure each. (They would not say that these were independent measurements). They stated that the team works from 7 AM to 7 PM, but would not say how many days a week they work at this pace. They stated that the average rate of measurement was 400 feddans of cotton per day (summer) and 300 feddans for wheat (winter). When questioned about these, they insisted that these were the correct rates. They use a casaba (a 3.55 meter stick) to do most of their measurements.

A visit to the headquarters in Cairo gave a more realistic account of their work. In agriculture they do area estimation for cotton, wheat and rice. They also do cropping pattern estimation for the MPWWR. They have 15,000 employees down to the district level. District offices may have from 30 to 70 people per office. They use 1-2500 scale maps with current features marked on them. The cadastral maps show all parcels of land. The staff plot fields on the map and send the maps to Cairo where exact areas are measured. Field staff use the 3.55 meter casaba for most measurements.

ESA's area measurements are not influenced by any government officials, and thus are said to be independent. However, much of their field work is done in collaboration with the extension agents and, thus, there may be some dependencies at that level. It takes 1 1.5 months to do the field work and send the maps to Cairo. ESA does take into account the non-productive areas within the hodes. Interplanted areas are broken out. Daily work capacity for the team is about 100-150 feddans. They do a 50% check of the team's work at the district level and a 25% check from the national level.

There are some differences in area estimates between MALR and ESA due to different methods. The ESA sample is set and used for 5 years. It is based on hode maps, which are summed to the village and the district levels. ESA staff use direct measurements and do not

get any external influences. MALR area estimates are based on village areas which are summed to the district level. When a difference greater than 5% occurs between MALR and ESA estimates, the Sampling Office breaks down their areas, hode by hode. They find the hodes with big differences and choose a subsample of farmers to measure. They determine the correct data and that is used in the official estimate. Each agency believes their data to be the correct data.

The biggest problem in ESA's work is the scale and age of their maps. The maps are 1-2500 scale maps, but they do have many features on them, so they are useable. Before the team goes to the field any new installations are marked on the maps. Maps are printed out yearly and used only for one year. ESA does have some newer technology, but it does not lend itself well to area estimation work. GIS technology work may be of some help if a survey database can be tied to it.

3.2.4 Sampling General Directorate

Egypt started applying crop cutting techniques in 1955 to estimate cotton production and yields.

The Sampling Offices in each governorate conducts crop cutting, objective yield and research work requiring sample plots in farmer's fields. Observations are made and measurements are taken in these plots. The attempt is to be objective in data gathering and survey analysis so that results will be representative of the true values in the population (see table 3.1).

Crop Cutting Surveys. Crop cutting surveys are used to determine crop yield estimates. *The sampling frames used for the survey vary by governorate.* For Behira they use the MPWWR maps giving areas by hode. In Assuit they use a hode by hode listing of cultivated and non-cultivated areas from the Ministry of Security Record Number 15. All of the sampling frame materials are very old, many of inconvenient scale and most out-of-date. Names and characteristic of land and villages have changed. The Sampling offices have tried to maintain and update the materials as best they can.

The national headquarters determines the number of crop cutting samples for each governorate and crop. These sample sizes are based on analysis of the previous year's data. A few governorate offices are developing the capability to compute variances and sample sizes.

Table 3.1: Sample Sizes of Commodities for Which Crop Cutting Surveys are Conducted

Crop *	Plot Size (in meters)	Sample Sizes			
		Behira	Dakahlia	Beni Suef	Assuit
Barley	2 X 2	22		4	10
Canola	2 X 2	12	6		
Chickpeas	3 X 3.5	4			24
Cotton	3 X 3.5	430	340	250	250
Fava Beans	3 X 3.5	124	125	62	74
Maize 1/	3 X 3.5	236	200	180	180
Onions	3 X 3.5	38	63	88	50
Peanuts	3 X 3.5	61		8	24
Potato	3 X 3.5	436	232	20	20
Rice 1/	2 X 2	470	480		
Sesame	3 X 3.5	14		8	28
Soybean 1/	3 X 3.5	10	14	56	24
Sugar beets	3 X 3.5	30	90	8	
Sunflower	3 X 3.5	32	14		8
Wheat 1/	2 X 2	450	450	280	350
Lentils	6 x 7		6		26
Sorghum	3 x 3.5			20	186

* Pilot work was tried for citrus, grapes, and lentils, but was discontinued.

1/ They used to take a sample of grain to dry for moisture determination. Now just use standard factors to adjust.

A stratified multistage sampling procedure is followed to select samples. The land areas are classified into strata based on type of irrigation and age of tile drainage. Groupings of similar land areas are formed into clusters. *The cluster sizes vary depending on the governorate.* For example in Behira the cluster size is 150-200 feddans, Dakahlia about 200-250 and Assuit 300-500 feddans. Clusters for use in the surveys are selected by systematic random sampling. Then sampling staff list and summarize all crop area in the selected clusters.

Sampling units are formed within the selected clusters consisting of about 3 feddans. If a prospective sampling unit exceeds 5 or 6 feddans, it is divided by 3 to determine the number of sampling units which will represent the area. A random sample of two sampling units is selected from each cluster for the crop cutting survey.

The selected farmer is notified by the sampling office or extension agent to wait until the sampling team arrives before harvesting his field. Farmers do not always follow these requests and some fields are lost.

The sampling team draws a map of the shape and dimensions of the selected field. The dimensions of the sample plot are subtracted from the field length and width. Random numbers less than or equal to these adjusted measurements are selected from a random number table. The location of the sample plot is determined by starting from the southwest corner of the field and measuring the random number of meters along and into the field. A stake is placed at the starting point of the unit. A surveyor's triangle is used to assist in getting right angles while laying out the plot. Stakes are put at each corner of the plot and a string is stretched around the corner stakes. The plot diagonals are measured as a check on the lay out and size. For row crops one dimension of the plot is determined by the location of the last parallel row within the plot dimension. The plot measurement is taken at the location of the last row and an adjustment of the plot size determined for later computation.

The farmer usually harvests the sample plot for the sampling staff person, who then weighs the grain and returns it to the farmer. The map, measurements and weights are then taken to the office where they are recorded on forms. Village, cluster number, name of farmer, random number, date, weight of grain in plot, and correction factor are recorded, then corrected measurement, average weight per sample and estimated weight per feddan are computed. Adjustment to local harvest measurement units may also be necessary. Weighted averages are computed for kilograms per feddan using the corresponding crop areas for districts.

During the summarization process an adjustment is made for *net productive area*. This is a reduction in total land due to canals, service roads and other non-productive land uses. Each district has its own service coefficients to adjust the total land to net productive land. These coefficients are apparently well understood, but no information was given as to how often the coefficients are updated or their precision. It is not clear why this additional adjustment is needed since actual field areas are used in the area estimation.

There are several potential procedural problems observed here:

- With small plot sizes, there is a need to be very precise in laying out the plot and harvesting the grain. The improper inclusion or exclusion of one stalk of grain in the

sample plot has serious consequences. In a 3 X 3.5 meter plot each grain head represents 400 others in that feddan; in a 2 X 2 meter plot each grain head represents 1050 others. Proper training and equipment can greatly reduce these errors.

- Harvest loss can sometimes be a problem particularly during mechanical harvesting. Grain dropped on the ground or run over by the harvesting machine is lost from harvested production. (Eventhough it may be utilized by grazing animals or produce volunteer crop in the next season, it is still lost from harvest production). The sampling staff has done some harvest loss research, but is currently using only technical coefficients to adjust for harvest loss. They felt that farmers took care while harvesting and little crop was lost.
- In some countries it is necessary to measure moisture content in harvested grain to determine quality, storage capability and true weight for market transactions. These measurements should be considered in Egypt also.
- Problems found to be common to all governorate sampling offices:
- Old equipment: some does not function well and is often heavy to carry and hard to use. Errors are sometimes caused by poor equipment. Need modern, smaller and lighter equipment.
- Transportation: some villages are hard to reach because the sampling offices have few cars; most motorcycles are 20 years old and in need of repair or replacement. Many staff are paying for repairs out of their own pockets.
- Communications between offices, villages and farmers are difficult.
- Timing: It is hard to determine the right time to visit the farmer. The offices have no authority over farmers or extension agents, so farmers may harvest before they get there. One office was able to minimize sample loss by having its staff check often in the village and find out when the harvest was to take place.
- Old sampling frame maps and listings.
- Lack of proper training: Many of the staff are nearing retirement age, and no new staff are available to learn and continue the work.
- Office space is very limited, so working conditions are difficult.
- Low pay: “good will” only lasts a short while.

Some additional noteworthy observations:

- A reward is given to governorate officials based on their crop yields. This puts a lot of pressure on leaders to influence their yield figures. [One field surveyor commented that he had to watch local officials who like to observe the crop cutting work, to

prevent them from throwing a little extra grain into the sample harvest]. We heard many comments that statistics staff were under pressure to improve yield even if it meant reducing area or increasing production figures to unrealistic levels. Rewards for falsifying information are always counterproductive!

- The team finds that there is lack of cooperation between Agricultural Affairs staff and sampling staff even within the same location.
- When sampling staff go to apply their yield estimate to the crop area estimates, they often found that these area estimates have been changed, particularly for cotton and rice. Remember that the number and location of samples were determined based on the initial area estimates of the Agricultural Affairs and ESA figures. Any changes in area without good evidence for change could cause serious credibility and quality questions. Some expressed feelings that these changes were being made due to external influences for economic or political reasons.

Objective Yield Surveys. In the 1980s, through a major USAID project aimed at improving data collection, objective yield techniques were tried for main field crops such as cotton, rice, maize, wheat and potatoes, and fruit like oranges and grapes. Objective yield techniques have the advantage of being able to forecast crop yields three to five months before harvest begins, while crop cutting techniques only estimate yield once harvesting has begun.

Objective yield techniques have been proven to be very beneficial for early warning of impending shortages, or the occasional surplus. Governments and traders with prior knowledge of shortage or surplus can go to international markets to meet their needs before prices go up.

The work was reasonably successful, but ran into some operational problems mostly related to administration of the work. Interestingly, Agricultural Economic Research Institute (AERI) is continuing to do some objective yield research for cotton and wheat, but their results are only published in research papers and given to the Director of the Agricultural Research Center.

4. METHODOLOGY

The methodology adopted in this study, to assess the quality of agricultural data produced by the MALR, followed three main dimensions:

1. Investigating and assessing the performance of the original sources of agricultural data. This included identifying the data available, problems and constraints that face these source organizations and how they impacted data quality. These tasks were accomplished by conducting institutional and statistical analyses on data produced by the MALR. A database of published MALR data was created relating to production and costs of production for the major field crops at governorate level (compiling existing data in electronic form).
2. Diagnosing the paths and procedures of the data from initial collection through to publication. This included review of procedures at all of the intermediate offices involved in the data collection, tabulation and publication process. Further, to evaluate the adopted methodology and techniques applied in collecting data. It also includes examining the relationships between the headquarters office and the offices at the various levels in the selected governorates.
3. Use time series techniques to analyze and assess the published MALR series of agricultural data. Published area, yield, production and costs of production data for major field crops were studied. The purpose was to see if these techniques gave insight to data quality and usefulness.

4.1 Investigation of Statistical Offices

The procedures followed to examine and analyze institutional performance of the MALR organizations participating in data collection and publication relied mainly on conducting surveys within four governorates. These governorates, Dakahlia, Behira, Beni-Suef, and Assuit (two in Lower, one in Middle, and one in Upper Egypt) were chosen to give broad geographical coverage. The main objectives of these surveys were to identify data available at different administrative levels, determine data accuracy and quality, and investigate possible sources and methods for collecting unavailable data.

4.1.1 Identifying Data Available at Different Administrative Levels

The team members visited two villages in each of three districts within the selected governorates. They were orienting themselves and gathering facts, in order to design the survey. The main survey work was done by PhD researchers, each assigned to gather required information from their respective governorates. The surveys were done in two villages within each of three districts. A major job of these researchers was to determine the data available at each administrative level and to gather the requested data. If it was not possible to obtain the requested data, then an explanation was necessary from the researchers.

The team designed a questionnaire for the interviews and a form to collect data at different

administrative levels. Various organizations were interviewed and examined such as: Statistics Office at both governorate and district levels, Sampling Office at governorate level, Egyptian Survey Authority (ESA), Village Cooperative Office (including extension agents and the chairman of the cooperative).

Throughout all of their work, researchers were advised to be alert to factors that affect data quality (completeness, precision, coverage, timeliness, etc.) and for any indications of influence or possible changes to the true values. However, in interviewing the official people at each administrative level, it was recommended that the researchers stress that the purpose of the survey was to determine the data availability, getting specific data to demonstrate that it was available and for use in some analysis.

The MVE team created databases from the MALR statistics to verify what data was available and to provide data for the time series analysis. The development process also helped identified problems in obtaining and verifying data while putting it in electronic media. The advantage of having databases is their efficient use in analysis.

Other agricultural sector data sources (CAPMAS, MTS, GTZ, IFPRI) were investigated so that comparisons with MALR data sources could be made and data quality assessed.

4.1.2 Determining Data Accuracy and Quality

The data gathered during team and researcher field trips were very valuable in the determination of data quality. Visual and computer checks of data within each office were made. The data as it traveled from one office to the next higher office was followed as far as possible. Influences on data, both internal and external, were investigated. Problems and constraints on the offices and the statistical process were identified. All of these factors affect the accuracy of data.

Statistical analysis on each set of data included plots of data, identification of outliers, variance computations, paired difference tests, and analysis of variance. Edit checks for consistency and reasonable relationships identified some unusual values, which required further verification.

4.1.3 Investigating Possible Sources and Methods for Collecting Unavailable Data

One of the important aspects of the visits to statistical offices was to determine how the data needed for a modern economy could be obtained. Researchers had questionnaires designed to collect data not currently available from statistical sources. They interviewed three farmers in each village to test ways to ask questions and the farmers willingness to respond.

4.2 Implementation of Field Investigations

In order to diagnose the paths and procedures of the data from initial collection through to publication it is necessary to visit each office involved.

The following sections include the major offices examined and the main activities which were implemented at each level. Annex E contains all of the details.

4.2.1 MALR Statistics Office, Governorate Level

The researchers were asked to determine the agricultural statistics that were available from the governorate office. A list of possible statistics was used as a guide. However, researchers were to probe for other statistics that the office might have or for other organizations which might be gathering statistics.

In each governorate the researcher needed to obtain a copy of the district summary (listing of district totals added to governorate total) for the specified crops. These usually included cotton, wheat, maize, rice (fava beans in Beni Suef), potatoes, a major fruit crop, cattle and sheep.

It was suggested that the team start with the most current year and work back through the years as far as they could until 1990. At least five years of data were needed. However, the office might not have all of the data items. If it did not, the researchers were told to make a note and give the reason (i.e., office does not collect, lost, eaten by mice, etc.).

Throughout all of the work, the researchers were asked to keep alert to factors affecting data quality and possible sources of potential errors.

4.2.2 Egyptian Survey Authority Office, Governorate Level

The researchers were assigned to look at the area measurement data for the selected villages, and select a representative *hode* in each village. For the selected *hodes* they had to get a copy of their measured areas for three years (if possible). The intention was to compare their hode area with those of the village data.

4.2.3 MALR Statistics Office, District Level

The researchers had to determine the agricultural statistics that were available from the office. They used the list of possible statistics as a guide in asking and recording the information. However, probing was recommended to obtain other statistics that they might have or for other organizations which might be gathering statistics.

In each district the researchers had to obtain a copy of the village summary (listing of village totals added to district total) for the specified crops and for years back to 1990 (as far back as possible). These usually included cotton, wheat, maize, rice (fava beans in Beni Suef), potatoes, a fruit crop, cattle and sheep.

This information should be what the district sent to the governorate office and in most cases would be exactly the same. If not, they probed to see if there was a known reason, but did not change the data. It was suggested that the researchers start with the most current year and work back through the years as far as they could. It was hoped to get at least five years of data. However, the office might not have all of the data items. If it did not, they made a note and gave the reason (e.g., do not collect, lost, eaten by mice).

4.2.4 MALR Village Cooperative Office

Again the researchers were asked to determine the agricultural statistics that were available from the office. They were advised to use the list of possible statistics as a guide in asking about and recording the information. However, probing for other statistics that they might have or other organizations which might be gathering statistics was suggested.

A recording form was supplied for each crop and livestock item. The form had a place to record the area, yield, production, cost of production and farmgate price. Two versions of the form were provided, depending on how the cost of production could be provided (by input or by operation). Livestock forms had a place to record number of head by age category and gender. A few cost of production items were suggested.

In each village the researcher was asked to gather data for the same items as were gathered at the district level. This information should be what the village sent to the district office and in most cases would be exactly the same. If it was not, the researchers were to probe to see if there was a known reason for the differences.

The researchers were assigned to review the data obtained from the Egyptian Survey Authority office for the selected hode. They had to choose five fields that were measured and then recorded their area on the form. The extension agent's area for the same field were recorded it in the appropriate column. Similar procedures were followed to fill in the form for the three years for which they had measurement data.

In each village the researchers were asked to interview at least three farmers of different size operations. They were trained to start with general questions about his operation (i.e. area, yield, production, number of livestock), then move into questions about labor times and costs to perform different farming operations, prices received for his crops and paid for inputs. The purpose was to determine the farmer's ability to recall such information for current and the past few years and also to determine if he was willing to reveal what stocks he kept for family use, and other information about his family. Finally the researchers were asked to write a short summary of the interview with each farmer for review by the team at a later date.

4.3 Time Series Analysis

4.3.1 Introduction

The reader is probably familiar with single-equation regression and perhaps with simultaneous-equation regression models. With regression we try to identify and describe the relationship between independent variables (cause) and dependent variables (effect). For example, we might try and estimate cotton production by measuring the amount of seed planted, fertilizer applied, pesticides applied and amount of irrigation applied. One could run many experiments and record these independent variable values along with their corresponding production. The regression analysis would result in an equation, which would model (describe) the relationship between these variables. If this relationship is valid, then one should be able to measure these independent variables for a field and apply the model parameters and arrive at the production of the field.

One problem with regression analysis is that the causal variables are not independent nor stationary. In 1976 the Box-Jenkins method was introduced. It analyzes the probabilistic

properties of economic time series on their own or “let the data speak for themselves.” This method goes through many mathematical operation to arrive at the time series (TS) model. An overview of the process is given in the next section.

4.3.2 Overview of Time Series Analysis

This section gives a brief overview of the process of time series studies.

Time series studies usually contain two aspects: analysis and modeling. The *analysis* process seeks to identify the statistical properties and characteristics of the series of data, in order to gain insight as to what kind of formal model might be appropriate. Graphical plots of the original data, autocorrelation functions and partial autocorrelation functions are major tools during the analysis phase. These tools help decide the degree of homogeneity in the time series, i.e., how many times the time series must be differenced before a stationary series results. The decision is made by looking at the autocorrelation functions for the time series and its differences. After the degree of homogeneity has been specified, the orders of the moving average and the autoregressive parts of the model must be determined.

The *modeling* process begins with the specification of the model and estimation of its parameters. After the model parameters have been estimated, then one performs a diagnostic check on the model’s performance. This usually involves looking at the autocorrelation function of the residuals from the estimated model. Then one performs a chi-square test to determine whether or not the residuals are themselves uncorrelated.

If the model passes the diagnostic check, then one can evaluate and analyze the model’s capability to forecast accurately. In addition, one can perform an ex-post forecast, comparing the forecast to actual data and evaluating its performance. This kind of analysis can help the researcher decide how far into the future the model can be used for forecasting.

In this study, 182 time series models are constructed. The results of these analyses are included in a separate volume with a detailed explanation of the procedures followed.

4.3.3 Time Series Analysis in This Study

(TS) analysis is usually used for describing the behavior of a variable and then forecasting this variable. In this study the primary purpose of (TS) analysis is to analyze the patterns in the variables. The variables are the basic agricultural data of Egypt: area, yield, and production, plus the cost of production. If any abnormal patterns are found in these variables, the forecasting ability of (TS) analysis is also used to further explore the consistency or inconsistency of the patterns found. That is, the study does not forecast the variables under study as part of an economic analysis, as would often be the case.

The null hypothesis in this analysis is that the data on different crops are collected properly by field staff and sent through a chain of offices and reach their final destination as accurate data. One normally expects to find differences in the patterns of data across crops and geographical areas. One would expect these differences despite the highly irrigated nature of Egyptian agriculture, because there are still differences in micro-climates, in the amounts of

water available each year in different locations, and in the incidences of pests and diseases, to name just a few of the factors that influence the variables under study.

The analysis will reject the null hypothesis, leading one to conclude that there may be interference in the collection and/or processing of the data, if many or all the variables show the same basic patterns, as measured by the parameters of (TS) analysis. In addition, one would not expect forecasts of the variables to have extremely low levels of error. If such low levels of error are found, this would be an additional evidence of unusual consistency in the data.

5. FINDINGS: COLLECTION PROCEDURES

5.1 Results of Team Visits

This section discusses the observations and findings of several visits of the MVE team and its researchers to governorate, district and village offices. The first visits were for orientation and basic fact finding missions. Then a survey was planned, questionnaires designed and researchers sent to determine data availability, collection and handling procedures, and quality characteristics. The emphasis was to find actual procedures used to gather and process data. Annex D contains the detailed summary of the team visits.

With the diversity of geographic areas and agriculture it was expected that statistical procedures and operational problems would differ. It was surprising to find that most areas had common problems and constraints. This section will give a summary of findings and some observations.

5.1.1 Village Cooperatives

Villages ranged from 800 to 6,000 feddans of cultivated area. The cropping pattern varied by village. Land holders were primarily small owners, with about 60% holding less than one feddan, another 20-30% holding 1 to 3 feddans and about 10 % with over five feddans. The number of extension agents ranged from 3 to 20 based on the size of the village and intensity of cropping. Each agent was responsible for 100-400 feddans. Their primary work was to assist farmers with improved farming practices, pest control and other production related functions. They were paid incentives for promotion of selected programs.

All offices had some additional staff other than extension agents. About a third of the cooperatives offices had a telephone, most had reasonably stable electricity, few had hand calculators, none had training to do any statistical work, and all supplied their own office supplies. Most offices have maps on the wall color-coded showing how much of the land is in major crop usage. Most have maps of *proposed areas* in crops and *actual areas* in crops. When explaining why proposed areas are obtained, they said that a higher committee at the national level gives indicators of suggested levels of production to governorate offices which in turn pass the recommendations to district and village offices. These are recommendations, not requirements as in the past. Farmers are free to grow what they wish, but if any are undecided the agents can make suggestions.

Extension agents live and work in the villages on a daily basis. Some of them are farmers themselves. Agents do an excellent job of estimating areas in crops, especially for major crops like cotton, rice, maize. Cotton receives special emphasis. They have created maps or have each farmer's area listed on sheets or in books. They work daily with farmers and are thus aware of yields and production. Most feel they know yield and production levels. None take any objective measures. The nearest they come is to count bags of grain harvested. Counts of farm machinery and numbers of livestock are known by most agents, even though they are only required to report these every two years. They are very aware of farm costs of production and farmer prices received, as they discuss these with farmers during the season.

They are seldom asked to report these prices and costs to the district office. Farmers sometimes do not speak freely to strangers, but will confide in the extension agent who is often a member of his village.

All of the extension agents and their supervisors meet frequently at the cooperative office to discuss the village estimates and problems. Agents receive information about fertilizer, pesticide, and cultural practices and suggested cropping patterns at the meetings.

Much of the data available at the village is not forwarded upward through administrative levels, and other data is forwarded only as requested. The Village cooperative staff have very detailed information that could be used for sampling frames for economic and research purposes.

Regarding the area re-measurements by ESA, some extension agents never see the ESA staff when they come, while others stated that the ESA staff used their more detailed maps.

5.1.2 MALR District Level Offices

Districts contained 30 to 50 cooperatives or villages. The number of cooperatives with telephones ranged from 1 to 5 in the districts visited. Data is usually carried to district offices by cooperative heads. They have meetings once or twice a week at the district office and discuss new pest control, cultural practices and village estimates of area, yield and production. Some commented that village estimates are determined at these district meetings, "the estimates of total production is set in a meeting, summarized and sent to the governorate office."

Most District Statisticians worked without any staff. They supplied their own supplies and hand calculators. Most had no training. Annex B contains all the details of the survey.

5.1.3 Governorate Level Offices

We talked to Agricultural Sector Chairmen, Heads of Agricultural Affairs offices, Governorate Statisticians and Sampling offices. Most recognized the need for good agricultural databases and series of data.

Governorate Statisticians usually have a few staff members to assist them.

5.1.4 General Observations

- **No Clear Line of Authority or Responsibility and Lack of Technical Direction.**

In theory the national EAS/MALR directs the statistical program, but little direction is given as they have no authority to enforce compliance. Several statistical offices stated that they had no contact with their national counterparts. They did not even receive forms or instructions on how to gather the data that was needed nationally.

Many offices expressed appreciation for our visits. “No one had ever asked them about their work, or shown any interest in what they were doing.” We had some governorate officials travelling with us (sometimes national officials) and it was clear that the village and district staff had never seen or communicated with this level official before. These officials received new insight into how work was actually being conducted and conditions under which their staff worked.

- **Lack of Training**

Village extension agents have training sessions on pesticide and cultural practice improvement. None had any training or instructions on gathering and reporting data.

Most governorate and district staff had no educational background or training helpful to their work.

- **Lack of Communication**

A few villages had telephones; none had faxes or copiers. When extension agents needed rapid communication, they used their home phone or the phone of a neighboring farmer.

Most governorate offices had access to a phone, but the quality of the connection was usually poor.

A few governorate sampling offices had a fax machine and a computer, which were supplied by external projects.

- **Lack of Resources**

At every level the employees were doing jobs for which they were not trained, they were not given equipment and supplies needed to do their job, and very few received any incentive pay. The lack of support grew more critical as we moved down through administrative levels. For example, the village extension agents and cooperative chiefs were supplying their own record books, paper, pens, and paying for copies out of their own pockets. Most brought hand calculators from home when they needed to use one.

A few district and governorate statistical offices did have a hand calculator supplied, but it was passed around among six employees.

Storage of records was particularly critical. None of the offices had filing cabinets. Records were stored in desk drawers or cabinets wrapped in paper and tied with string. Mice gnawing on record packages often resulted in records not remaining intact for more than a year or two.

- **No Incentives to Do a Good Job**

Incentives can come in several forms. Of course, money is the most tangible. However, a word or letter of appreciation or encouragement can improve attitudes and performance. A token gift or symbol can lift one's spirit.

In the villages, extension agents get incentives for the implementation of improved horticultural, pesticide and management practices, but nothing for statistical work.

- **No Standardized Forms or Procedures**

Virtually every office made its own forms using the backs of old questionnaires, scrap paper or paper they purchased themselves. They used carbon paper to make copies. Imagine the condition and quality when using the third or fourth carbon copy.

5.2 Data Available and Their Relationship To Data Needs

5.2.1 Existing Data Vs Available Data

In Egypt there are a lot of data. In Table 5.1 note that some data go back to 1911. The problem is that one would have to go to the agricultural library and dust off a historic volume, it is not readily available. When trying to get some data, we have asked if it was available and were told that Mr. X had it. When we contacted Mr. X, he said it was locked in a file cabinet and he did not have the key that day, or it could only be released with the Minister's approval. This data apparently exists but is not available. Much of Egypt's intermediate summary data is still on hand listed summary sheets. Thirty years of data may exist on governorate and district summary sheets but is of little use for analysis. Readily available data is necessary for the multitude of analyses needed for the privatized economy.

If information is not disseminated, it is useless. There is little reason to expend the time, effort and funds to gather and process the data.

5.2.2 Agricultural Data Collected by MALR

One of the main publications of the EAS is the "Annual Bulletin of Agricultural Economics", the main source of official agricultural statistics in Egypt. This agricultural statistical yearbook contains, among others, data on the following:

- Components of Agricultural National Income.
- Production and consumption of food commodities (food balance sheet)
- Cropping pattern and land uses.
- Area and production of field crops, vegetables and fruits at the governorate level.
- Costs of production per feddan for field crops and some vegetables.
- Average farmgate prices for main crops.
- Monthly distribution of agricultural labor.
- Number of farm livestock at the governorate level.
- Production of meat and dairy products.

- Fish production.

Between 1989 and 1994 nothing was published. In 1995 they started publishing the missing data. All of it has now been published and the current publications are now back on schedule.

Table 5.1 “Existing Agricultural Data” identifies the area, yield and production of field crops, vegetables and fruit for which data are collected. The table also gives the when coverage began, level of coverage and source of the information. Some of the data series go back to 1911. However, little of the old data is of value except for historic purposes. Data prior to the latest agricultural privatization reforms do not represent current sector characteristics, and would be of little use in modeling. Another big problem is the accessibility, for to obtain much of the data would require a trip to the agricultural library or other obscure office.

Note that in Table 5.1 there is a source D for database information that has just been entered by this Data Quality effort. It consists of data for the four governorates in which we did our studies. Area, yield and production, cropping pattern, prices and cost of production data are included in different computer files. Once checked out it will be available for use in studies and analysis. The team also did time series analysis on much of the data. The results of the analysis are included in chapter 6 of this report.

5.2.3 Data Requirements

Table 5.2 represents the basic statistics of use to producers, and decisions makers, policy makers, and economic and impact analysts. It lists data that are needed for a complete study and understanding of agricultural sector economics. Decision makers at all levels of the sector need some of this information to carry out their business. This is not intended to be a complete listing of data needs, but does contain most items of interest. However, looking over the needed information, it is very clear that most of the data are not even available. It will be necessary to develop a statistical data collection, processing and dissemination system that can handle these new requirements.

Table 5.1: Existing Agricultural Data

Crop	Estimates			Begin	Level Of	Source	Comments
Field Crops	Area	Yield	Prod.	Coverage	Coverage		
Winter Crops -							
Wheat	X	X	X	1911	N,G,D	C,S,D	1)
Barley	X	X	X	1911	N,G,D	C,S	
Onion	X	X	X	1935	N,G,D	C,S	
Fava Beans green	X	X	X	1911	N,G,D	D	2)
Fava Beans - dry	X	X	X			C,S,D	2)
Garlic	X	X	X	1945	N,G,D	C,S	
Sugar Beets	X	X	X	1982	N,G,D	C,S	
Lentils	X	X	X	1912	N,G,D	C,S	
Chickpeas	X	X	X	1928	N,G,D	C,S	
Lupines	X	X	X				
Fenugreek	X	X	X				
Flax	X	X	X	1923	N,G,D	C	
Clover	X			1990	N,G,D	C	
Perm. Berseem	X			1939	N,G,D	C,D	2)
Temp. Berseem	X			1939	N,G,D	C,D	2)
Canola	X			1997	N,G,D	C,S	
Summer Crops -							
Cotton	X	X	X	1956	N,G,D	C,S,D	1)
Maize	X	X	X	1935	N,G,D	C,S,D	1)
Sorghum	X	X	X	1935	N,G,D	C,S	
Rice	X	X	X	1911	N,G,D	C,S,D	1)
Sugarcane	X	X	X	1911	N,G,D	C,S	
Soybean	X	X	X	1966	N,G,D	C,S	
Potatoes	X	X	X	1945	N,G,D	C,D	3)
Onion	X	X	X	1935	N,G,D	C,S	
Peanut	X	X	X	1912	N,G,D	C,S	
Sesame	X	X	X	1928	N,G,D	C,S	
Sunflower	X	X	X	1966	N,G,D	C,S	
Medical & Aromatic	X	X		1932	N,G,D	C	
Nili Crops -							
Maize	X	X	X	1935	N,G,D	C,D	1)
Rice	X	X	X	1911	N,G,D	C,D	2)
Sorghum	X	X	X	1936	N,G,D	C	
Potatoes	X	X	X	1945	N,G,D	C	
Onion	X	X	X	1935	N,G,D	C	
Feeds	X			1980	N,G,D	C	

Continued Table 5.1: Existing Agricultural Data

Fruits	Estimates			Begin	Level Of	Source	Comments
	Area	Yield	Prod.	Coverage	Coverage		
Orange	X	X	X	1951	N,G,D*	C,D	4)
Balady	X	X	X	1951	N,G,D*	C	
Sucari	X	X	X	1951	N,G,D*	C	
Valencia	X	X	X	1951	N,G,D*	C	
Abasora	X	X	X	1951	N,G,D*	C	
Kalili	X	X	X	1951	N,G,D*	C	
Abudomu	X	X	X	1951	N,G,D*	C	
Yahfaywy	X	X	X	1951	N,G,D*	C	
Youssefi Soleman	X	X	X	1951	N,G,D*	C	
Youssefi Lemon	X	X	X	1951	N,G,D*	C	
Sweet Lemon	X	X	X	1951	N,G,D*	C	
Lime	X	X	X	1951	N,G,D*	C	
Mandarin	X	X	X	1951	N,G,D*	C	
Sour Orange	X	X	X	1951	N,G,D*	C	
Grapefruit	X	X	X	1951	N,G,D*	C	
Grapes	X	X	X	1951	N,G,D*	C,D	4)
Pear	X	X	X	1951	N,G,D*	C	
Mango	X	X	X	1951	N,G,D*	C	
Apricot	X	X	X	1951	N,G,D*	C	
Banana	X	X	X	1951	N,G,D*	C	
Olive	X	X	X	1951	N,G,D*	C	
Pomegranate	X	X	X	1951	N,G,D*	C	
Guava	X	X	X	1951	N,G,D*	C	
Prickly Pear	X	X	X	1951	N,G,D*	C	
Fig	X	X	X	1951	N,G,D*	C	
Apple	X	X	X	1951	N,G,D*	C	
Plum	X	X	X	1951	N,G,D*	C	
Kaki Maslimela	X	X	X	1951	N,G,D*	C	
Pecan	X	X	X	1951	N,G,D*	C	
Chiskda	X	X	X	1951	N,G,D*	C	
Blackberry	X	X	X	1951	N,G,D*	C	
Tomatoes						D	93-97

Continued Table 5.1: Existing Agricultural Data

Vegetables	Estimates			Season			Begin	Level Of	Source	Comments
	Area	Yield	Prod.	Winter	Summer	Nili	Coverage	Coverage		
Tomato	X	X	X	X	X	X	1952	N,G,D	C	
Squash	X	X	X	X	X	X	1952	N,G,D	C	
Green Beans	X	X	X	X	X	X	1952	N,G,D	C	
Dry Beans	X	X	X	X	X	X	1952	N,G,D	C	
Kidney Beans Green	X	X	X	X	X		1952	N,G,D	C	
Kidney Beans Dry	X	X	X	X	X		1952	N,G,D	C	
French Beans Green	X	X	X	X	X	X	1952	N,G,D	C	
French Beans Dry	X	X	X		X	X	1952	N,G,D	C	
Broad Bean	X	X	X	X						
Egyptian Melon	X	X	X	X			1952	N,G,D	C	
Beet	X	X	X	X			1952	N,G,D	C	
Cabbage	X	X	X	X	X	X	1952	N,G,D	C	
Cauliflower	X	X	X	X	X	X	1952	N,G,D	C	
Eggplant	X	X	X	X	X	X	1952	N,G,D	C	
Green Pepper	X	X	X	X	X	X	1952	N,G,D	C	
Okra	X	X	X		X	X	1952	N,G,D	C	
Jews Melon	X	X	X	X	X	X	1952	N,G,D	C	
Spinach	X	X	X	X		X	1952	N,G,D	C	
Sweet Potato	X	X	X	X	X	X	1952	N,G,D	C	
Taro or Dasheen	X	X	X	X	X	X	1952	N,G,D	C	
White Radish	X	X	X	X	X	X	1952	N,G,D	C	
Turnip	X	X	X	X	X	X	1952	N,G,D	C	
Lettuce	X	X	X	X	X	X	1952	N,G,D	C	
Carrot	X	X	X	X	X	X	1952	N,G,D	C	
Parsley	X	X	X	X	X	X	1952	N,G,D	C	
Rocket	X	X	X	X	X	X	1952	N,G,D	C	
Egyptian Leek	X	X	X	X	X	X	1952	N,G,D	C	
Leek	X	X	X	X	X	X	1952	N,G,D	C	
Dill	X	X	X	X	X	X	1952	N,G,D	C	
Coriander	X	X	X	X	X	X	1952	N,G,D	C	
Chard	X	X	X	X	X	X	1952	N,G,D	C	
Strawberry	X	X	X	X	X		1952	N,G,D	C	
Artichoke	X	X	X	X		X	1952	N,G,D	C	
Chichory	X	X	X	X			1952	N,G,D	C	
Watermelon	X	X	X	X	X	X	1951	N,G,D	C	
Amer. Watermelon	X	X	X		X	X	1951	N,G,D	C	
Melon	X	X	X		X	X	1951	N,G,D	C	
Cucumber	X	X	X	X	X	X	1951	N,G,D	C	
Snake Cucumber	X	X	X	X	X	X	1951	N,G,D	C	
Melon (Quoz Asal)	X	X	X		X		1951	N,G,D	C	
Potatoes *	X	X	X	X	X	X				

Level of Coverage

N = Arab Republic of Egypt

G = Governorate

D = District

V = Village

Source

P = Published Information

D = Database

S = Sampling

C = Current Ag. Statistics Office

N.B. : * = District since 1980

Behira, Dakahlia, Beni Suef, Assuit

1) 70-79 Gov. totals 80-96 Dist. totals

2) 80-97

3) 80-92

4) even years 82-94

Table 5.2: Statistics of Use to Producers, Processors, Traders, Decision Makers, Policy Makers, and for Economic and Impact Analysts

Potential Data Items	
Land Utilization	Imports - volume, price,
Total	by quality class
Arable, Cropland	by commodity
Irrigated	
Non-irrigated	Exports - volume, price,
Pasture, Grazing	by quality class
Forest, woods	by commodity
Fallow	
	Consumer Income and Expenses
Land Tenure	Rural and Urban
Owned, Rented Out, Rented In	Capital Inputs
Share Cropped	Farm Credit Needs
Crop Area	
Planted, to each crop	Employment in Agriculture
Harvested	Number, amount, cost, value
Crop Yield	On-farm, off-farm
	Paid, non-paid - family
Crop Production - amount, value	Paid, non-paid - hired
Planned (expected)	Labor - more detailed analysis
Actual	and other factors
Forecast	
Early Forecasts (Early Warning)	

Distribution of Crop Production-	Agricultural Facilities/Machinery Value of buildings and
Sold, Share, In-kind,	Other structures
On Farm, Seed , Storage	Farm equipment
losses (pre- and post-harvest)	Number, Usage and costs
	Tractors
Crop Damage Data	Buildings
Cropping Patterns for Major Crops	Storage Facilities
Growth Rates (Area, Prod., Yield, Plants)	Agricultural Cultural Practices
Crop Inputs - total, value	Marketed Crops
Fertilizer Use by type	Usage and Surplus
Pesticide Use by type	Storage facilities number, capacity
Seed Use by type	Type, characteristics,
Irrigation Use by type	Costs to store
Water Use	
Energy	Marketing factors and Costs
Fuel Use by type	Transportation, Fuel
Electricity Use	Storage and Shrinkage factors
	Export or Import Fees or Taxes
Crop Input - per hectare, per farm	Service Charges (Loading, Weighing,
Requirements vs Availability	Road Guard, Borderguard)
Type, Amount, Cost	
Rate of Application	Prices
	Farm Gate
	Wholesale

Livestock and Poultry	
Species, number, age, sex & use	Retail
Nomadic, Farm, Urban	Index Numbers
Births, Deaths, Slaughter	Index - Parity between prices received
Bought & Sold - Number and Price	and prices paid by farmers
Feed - Stocks and Storage	
Livestock Products	Farm Income
Beehives and production	
Fishery Statistics	Farm Costs and Returns
	Farm Budgets (spent and received)
Costs of Production	
Inputs (above)	Demographic Characteristics
Labor	Rural Population - age, sex, education
Harvest - equipment use, fuel	Health
Feed Costs	Economically active and inactive
Labor and Capital Inputs	
Health, Disease,	Reliability Statement of Estimates
Veterinary Services	

6. FINDINGS: TIME-SERIES ANALYSES

6.1 Introduction

Time Series (TS) analysis is used in many sectors of the economy. The main advantages of (TS) analysis are its ability to model (explain) the change of a variable through time, and to forecast the future values of the variable.

If a variable under study does not change over time, i.e., is stationary, it is essentially a constant and does not really need to be studied. However, most variables of interest do change over time, i.e., are nonstationary.

The type of nonstationarity must be identified and the data transformed to make it stationary. To proceed with the analysis, the immediate problem when considering nonstationarity is that there are many ways in which a process could be nonstationary. For example we have:

- Nonstationarity in the mean,
- Nonstationarity in the variance,
- Nonstationarity both in the mean and variance, etc.

The above three cases are illustrated in Figures (6-1) through (6-3).

The study of plots and of ACF and PACF correlograms on the original data are the procedures used to determine whether a (TS) is stationary or not, and further to determine the type of non-stationarity. The results of these plots then indicate the model that might best describe the time series data.

A general model capable of representing a wide class of nonstationary time series is the "Auto-Regressive Integrated Moving Average model (p,d,q)," abbreviated ARIMA(p,d,q). This class includes an unlimited number of models.

In this study the primary purpose of (TS) analysis is to analyze the patterns in the variables. If any abnormal patterns are found in these variables, the forecasting ability of (TS) analysis is also used to further explore the consistency or inconsistency of the patterns found. That is, the study does not forecast the variables under study as part of an economic analysis, as would often be the case.

The null hypothesis in this analysis is that the data on different crops are collected properly by field staff and sent through a chain of offices and reach their final destination as accurate data. One normally expects to find differences in the patterns of data across crops and geographical areas. One would expect these differences despite the highly irrigated nature of Egyptian agriculture, because there are still differences in micro-climates, in the amounts of water available each year in different locations, and in the incidences of pests and diseases, to name just a few of the factors that influence the variables under study.

Chart 6.1: Illustration of Time Series Data Showing Nonstationarity in the Mean

Chart 6.2: Illustration of Time Series Data Showing Nonstationarity in the Variance

Chart 6.3: Illustration of Time Series Data Showing Nonstationarity both in the in the Mean and Variance

The analysis will reject the null hypothesis, leading one to conclude that there may be interference in the collection and/or processing of the data, if many or all the variables show the same basic patterns, as measured by the parameters of (TS) analysis.

This chapter summarizes the results of 182 time series analyses. The study concentrated on time series data of area, yield, production, and total cost of production of the four main crops. As mentioned earlier the present study covers Behira, Dakahlia, Beni Suef and Assiut. Each governorate included three selected districts. The Time Series analyses were conducted at both the governorate and the district levels.

The selected governorates and districts and their corresponding codes are presented in Table 6-1.

Table 6.1: Code Numbers of Governorates and Districts

Gov. 2 El-Behira	Gov. 5 El-Dakahlia	Gov. 22 Beni Suef	Gov. 31 Assiut
Housh Aissa DS 6	El-Sinbilawain DS 5	Ahnasia DS 1	Abnoub DS 1
Damanhour DS 8	Sherbien DS 7	Beni-Souif DS 3	Abou-Tieg DS 2
Koom Hamada DS 12	El-Mansoura DS 10	El-Fashn DS 6	Assiut DS 3

Gov = Governorate

DS = District

Wheat, maize, and cotton are cultivated in all four governorates and their selected districts, while rice is cultivated in Behira and Dakahlya. Table 6-2 indicates the total number of time series cases analyzed in this study.

Table 6.2: Total Number of Time Series Cases by Governorate

Gov. & DS	Gov. 2 and its 3 districts	Gov. 5 and its 3 districts	Gov. 22 and its 3 districts	Gov. 31 and its 3 districts	Total
Variable					
Area	16	16	12	12	56
Yield	16	16	12	12	56
Production	16	16	12	12	56
Cost of Production*	4	4	3	3	14
Total	52	52	39	39	182

* The time series of the total costs of production are available only at the governorate level

The time series on the district level consists of annual data from 1980 to 1997, while the time series at the governorate level consists of annual data from 1970 to 1997.

6.2 Modeling Process with Illustrative Examples

To show how the time series analyses were carried out, we first discuss in some detail the construction of one of the 182 models. Consider the yield time series data for maize in Assiut Governorate (governorate level is denoted by DS 99). This time series consists of annual data from 1970 to 1997. The detailed steps used in the analysis are:

- [1] A plot of the original series, presented in Chart (6.4) below. The series itself rises steadily over time, indicating that it is non stationary in the mean
- [2] We differenced the original series twice to achieve stationarity [see Chart (6.5) below]
- [3] The autocorrelation function (ACF) is applied to the differenced series, shown in Chart (6.6). The result is a damped sinusoidal shape indicating a second-order autoregressive process.
- [4] The partial autocorrelation function (PACF), shown in Chart (6.7), has significant spikes at lags 1 and 2, confirming a second-order autoregressive AR(2) interpretation of the differenced series.
- [5] We thus estimate an AR(2) model, and the results are given below.

$$w_t = -0.8838 * w_{t-1} - 0.5703 * w_{t-2}$$

(-4.29) (-2.52)

Notice that the values within parentheses are the t-ratios. They indicate that the two parameters are significantly different from zero. Moreover the chi-square statistic is 11.8, which with 10 degrees of freedom (DF) is insignificant even at 95% significant level.

- [6] Finally the forecasts of the last two years together with their actual values are obtained. The errors of forecast are 3% and 5% respectively.

Final Estimates of Parameters of the example of Maize in Assiut Governorate Total (70-97) Yield (MW) Second Difference AR(2) are as follows:

Type	Estimate	St. Dev.	t-ratio
AR 1	-0.8838	0.2060	-4.29
AR 2	-0.5703	0.2266	-2.52

No. of Observations = 24

Residuals: SS = 37.3896 (back forecasts excluded)

MS = 1.6995 DF = 22

Modified Box-Pierce (Ljung-Box) chi-square statistic

Lag 12 24 36 48

Chi-square 11.8 (DF=10)

Forecasts from period 26

Period	Forecast	Actual
27	19.45	20.12
28	21.37	22.46

Chart 6.4: Original Series of Maize Yield in Assuit Governorate

Chart 6.5: Difference of Original Series of Maize Yield Data to Achieve Stationarity

Chart 6.6: The Autocorrelation Function of Maize Yield in Assuit Governorate During the Period 1970-1997

Chart 6.7: The Partial Autocorrelation Function of Maize Yield in Assuit Governorate During the Period 1970-1997

Each of the 182 time series models is constructed using similar techniques. The details of these are contained in 4 volumes:

Volume 1	Contains the 52 models of Behira Governorate
Volume 2	Contains the 52 models of Dakahlia Governorate
Volume 3	Contains the 39 models of Beni Suef Governorate
Volume 4	Contains the 39 models of Assiut Governorate

The following sections are a summary of the 182 time series models for the area, yield, production, and total cost of production of the four crops wheat, rice, maize, and cotton in the four governorates (Behira, Dakahlia, Beni-Souif, Assiut) and their twelve selected districts.

6.3 Area, Yield, Production, and Cost of Production in Behira

The summary results of the 52 estimated models are given in tables (6-3) through (6-15) below. The following notation is used in the tables:

F2	we difference the raw data twice to remove nonstationarity,
LN1	we make one logarithmic transform to remove nonstationarity
MA(2)	difference model of second degree for stationary series.
AR(1)	Autoregressive model of first degree for stationary series.

6.3.1 Wheat

Nine out of twelve of the time series of the area, yield, and production show obvious change in the mean i.e., these time-series are non-stationary in the mean. Therefore, we difference these nine time-series data once or at most twice to achieve stationarity. The other three time-series are non-stationary in both the mean and the variance. To convert them to stationary time-series, we used a logarithmic transformation followed by one or two difference transforms.

The appropriate models for the stationary series are (in most cases) either AR or MA models of at most third order. These identified models are accepted because the chi-square statistics are insignificant at 95% confidence level.

The forecast error does not exceed

6%	in <u>most</u> cases of the area models
2%	in <u>most</u> cases of the yield models
7%	in <u>all</u> cases of the production models.

6.3.2 Rice

All time series of the yield are non stationary both in the mean and variance. In order to remove non stationarity we first applied a logarithmic transformation and then differencing the logarithmic series once or twice at most. All the time series of the area and production (except one) are nonstationary in the mean. After difference these series twice at most we achieved stationarity.

The appropriate models for the stationary series are (in most cases) either AR or MA models of at most third order. The models are promising because they have the lowest chi-square statistics.

The forecast error does not exceed

- 10% in all cases of the area models
- 4% in all cases of the yield models
- 14% in all cases of the production model.

6.3.3 Maize

Nine out of twelve of the time series of the area, yield, and production show obvious change both in the mean and variance. Therefore, we first applied logarithmic transformation and then difference the logarithmic series once or twice to achieve stationarity. The other three time-series are nonstationary only in the mean. In order to remove nonstationarity from these three series we differenced them once or twice.

The appropriate models for the stationary series are as follows:

- The area models: 3 MA(1) models & one AR(1) model.
- The production models: all of them are AR models of at most 3rd order.
- The yield models: consist of a combination of AR, MA and ARMA models.

The forecast error does not exceed

- 4% in most cases of the area models.
- 3% in most cases of the yield models.
- 7% in most cases of the production models.

6.3.4 Cotton

The twelve time series of the yield, area, and production are non-stationary in the mean. Therefore, we difference these time series data once or twice to achieve stationarity.

Eleven appropriate models for the stationary series are either AR(1) or MA(1) models. The last identified model is ARMA(1,1). All of these models are accepted because they have insignificant chi-square statistics even at 95% confidence level.

The forecast error does not exceed

- 16% in all cases of the area models
- 15% in most cases of the yield models
- 21% in all cases of the production models.

6.3.5 The Time-Series Models of Total Cost of Production

The total cost time series of wheat, rice, maize, and cotton are positively trended. It means that these series are nonstationary in the mean. They are converted to stationary time series by differencing the raw time series data once in all four cases.

The four appropriate models for the stationary series of either AR or MA models of at most second order. These models seems most promising because they have the lowest chi-square statistics.

The forecast errors are as follows:

Wheat	2%
Rice	11%
Maize	35%
Cotton	6%

Table 6.3: Area of Wheat in Behira Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>MA(1)-F2</u> $w_t = a_t - 0.9751 * a_{t-1}$ (13.9)	9.5	206540 227374	203834 212977
Housh Aissa	<u>ARMA(3,2)-LN+F2</u> $w_t = -0.6933 * w_{t-1} - 0.8288 * w_{t-2}$ (-1.88) (-2.40) $-0.2857 w_{t-3} + a_t$ (-0.83) $-0.4860 a_{t-1} + 0.8499 * a_{t-2}$ (1.29) (2.82)	8.1	8209	9522
Damanhour	<u>MA(1)-F2</u> $w_t = a_t - 0.9376 * a_{t-1}$ (5.84)	8.6	26510	25118
Koom Hamada	<u>ARMA(1,3)-LN+F2</u> $w_t = -0.9311 * w_{t-1} + a_t + 0.251 a_{t-1}$ (-6.03) (1.18) $-0.5344 * a_{t-2} - 0.9499 * a_{t-3}$ (2.52) (4.22)	6.2	14444	14455

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.4: Yield of Wheat in Behira Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>MA(2)-F1</u> $w_t = a_t + 0.1236a_{t-1} + 0.4113*a_{t-2}$ (0.63) (2.10)	9.8	17.04 17.02	16.83 17.27
Housh Aissa	<u>AR(3) F2</u> $w_t = -0.8757*w_{t-1} - 0.8321*w_{t-2}$ (-3.77) (-3.71) $-0.7989*w_{t-3}$ (-3.53)	6.3	16.28	18.84
Damanhour	<u>AR(1)-F1</u> $w_t = 0.5419*w_{t-1}$ (2.50)	14.9	17.20	16.82
Koom Hamada	<u>MA(1)-F2</u> $w_t = a_t + 0.923*a_{t-1}$ (5.43)	10.5	18.22	18.05

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.5: Production of Wheat in Behira Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>AR(2)-F1</u> $w_t = 0.1741w_{t-1} + 0.4738*w_{t-2}$ (0.92) (2.49)	13.7	3519442 3869698	3436866 3662571
Housh Aissa	<u>AR(3) -F1</u> $w_t = -0.06w_{t-1} + 0.235w_{t-2}$ (-0.26) (0.95) $+0.617*w_{t-3}$ (2.3)	8.3	145984	144345
Damanhour	<u>ARMA(1,1)-F1</u> $w_t = 0.7911w_{t-1} + a_t$ (1.38) $-0.6357*a_{t-1}$ (1.88)	7.1	455918	422823
Koom Hamada	<u>ARMA(2,1)-LN+F1</u> $w_t = -0.903*w_{t-1} + 0.0957w_{t-2}$ (-2.73) (0.32) $+a_t + 0.9742*a_{t-1}$ (3.29)	6.9	263315	265004

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.6: Area of Rice in Behira Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>MA(1)-LN+F2</u> $w_t = a_t - 0.8653 * a_{t-1}$ (5.70)	9.5	212259 244698	213416 215561
Housh Aissa	<u>MA(1) -F2</u> $w_t = a_t - 0.945 * a_{t-1}$ (6.08)	16.2	3653	3631
Damanhour	<u>ARMA(2,1)-LN+F2</u> $w_t = -0.93 * w_{t-1} - 0.53 w_{t-2}$ (-2.22) (-1.10) + $a_t - 0.85 * a_{t-1}$ (2.75)	9	30944	31101
Koom Hamada	<u>AR (4)- F1</u> $w_t = -0.80 * w_{t-1} - 0.51 w_{t-2}$ (-2.75) (-1.87) + $0.80 * w_{t-3} - 0.90 * w_{t-4}$ (3.23) (-3.01)	8.5	8493	9345

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.7: Yield of Rice in Behira Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>ARMA(4,1)-LN+F1</u> $w_t = -0.98*w_{t-1} - 0.30w_{t-2}$ (-4.09) (-0.11) $+0.45w_{t-3} + 0.50w_{t-4}$ (1.66) (2.37) $+ a_t + 0.77*a_{t-1}$ (4.16)	8.9	3.647 3.687	3.666 3.702
Housh Aissa	<u>MA(1) -LN+F2</u> $w_t = -a_t - 0.9638*a_{t-1}$ (6.63)	16.0	3.653	3.662
Damanhour	<u>AR (2)-LN+F2</u> $w_t = -0.98*w_{t-1} - 0.89*w_{t-2}$ (-5.96) (-4.91)	17.2	3.747	3.759
Koom Hamada	<u>AR (2)-LN+F2</u> $w_t = -0.81*w_{t-1} - 0.61*w_{t-2}$ (-3.6) (-2.69)	5.1	3.646	3.8

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.8: Production of Rice in Behira Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>MA(1)-F2</u> $w_t = a_t - 0.8556 * a_{t-2}$ (5.66)	5.6	774108 902202	805153 772860
Housh Aissa	<u>AR(3) -F2</u> $w_t = -0.65 * w_{t-1} + 0.03 w_{t-2}$ (-2.8) (0.11) $+ 0.71 * w_{t-3}$ (2.64)	6.3	31470	29912
Damanhour	<u>ARMA(3,1)-F1</u> $w_t = -0.18 w_{t-1} + 0.41 * w_{t-2}$ (-0.79) (2.45) $+ 0.78 * w_{t-3} + a_t$ (3.3) $- 0.75 * a_{t-1}$ (2.18)	8.7	116320	118436
Koom Hamada	<u>AR (1)-F2</u> $w_t = -0.6667 * w_{t-1}$ (-2.2)	13.5	59506	55198

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.9: Area of Maize in Behira Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	$\frac{MA(1)-LN+F2}{w_t = a_t - 0.93*a_{t-1}}$ (5.63)	6.8	156709 160521	155593 156843
Housh Aissa	$\frac{MA(1)-LN+F1}{w_t = a_t - 0.95*a_{t-1}}$ (5.20)	8.9	9635	9228
Damanhour	$\frac{MA(1)-F1}{w_t = a_t - 0.60*a_{t-1}}$ (2.65)	8.8	15438	15595
Koom Hamada	$\frac{AR(1)-LN+F2}{w_t = -0.59*w_{t-1}}$ (-2.72)	16.6	16776	21226

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.10: Yield of Maize in Behira Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>ARMA(1,3)-F2</u> $w_t = -0.90 * w_{t-1} + a_t + 0.33 a_{t-1}$ (-3.64) (1.33) $-0.70 * a_{t-2} - 0.55 * a_{t-3}$ (3.4) (2.58)	7.1	24.4 25.12	22.81 25.72
Housh Aissa	<u>MA(1) -LN+F2</u> $w_t = a_t - 0.88 * a_{t-2}$ (4.78)	15.5	21.89	21.93
Damanour	<u>AR (1)-LN+F1</u> $w_t = -0.58 * w_{t-1}$ (2.71)	7.8	22.58	23.22
Koom Hamada	<u>ARMA (2,1)-F1</u> $w_t = -0.77 * w_{t-1} + 0.24 w_{t-2}$ (1.98) (0.75) $+ a_t - 0.87 * a_{t-1}$ (2.37)	5.5	28.67	27.87

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.11: Production of Maize in Behira Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>AR (1)-LN+F2</u> $w_t = -0.62 * w_{t-1}$ (-3.47)	8.3	3608924 3916712	3674766 3656438
Housh Aissa	<u>AR(3) -LN+F2</u> $w_t = -0.89 * w_{t-1} - 0.76 * w_{t-2}$ (-3.29) (-2.46) $-0.40 w_{t-3}$ (-1.48)	5.6	122164	123689
Damanhour	<u>AR (2)-LN+F2</u> $w_t = -0.68 * w_{t-1} - 0.53 * w_{t-2}$ (-2.70) (-1.92)	5.7	373331	359188
Koom Hamada	<u>AR(1)-LN+F1</u> $w_t = -0.58 * w_{t-1}$ (-2.61)	12.5	480962	559613

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.12: Area of Cotton in Behira Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>MA(1)-F2</u> $w_t = a_t - 0.888*a_{t-1}$ (5.42)	7.4	146816 170370	147196 146675
Housh Aissa	<u>MA(1)-F2</u> $w_t = a_t - 0.8797*a_{t-1}$ (4.74)	7.7	6578	5532
Damanhour	<u>AR (1)-F2</u> $w_t = -0.8592*w_{t-1}$ (-3.90)	10.9	25905	24659
Koom Hamada	<u>MA (1)-F2</u> $w_t = a_t - 0.87*a_{t-1}$ (4.47)	5.3	3537	3201

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.13: Yield of Cotton in Behira Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>AR(1)-F2</u> $w_t = -0.73 * w_{t-1}$ (-3.63)	13.3	6.46 6.52	6.57 5.03
Housh Aissa	<u>MA(1)-F2</u> $w_t = a_t - 0.93 * a_{t-1}$ (5.02)	7.5	7.14	5.34
Damanhour	<u>MA (1)-F2</u> $w_t = a_t - 0.86 * a_{t-1}$ (3.71)	8.6	6.33	6.04
Koom Hamada	<u>ARMA (1,1)-F2</u> $w_t = -0.91 * w_{t-1} + a_t$ (6.35) $+ 0.60 * a_{t-1}$ (2.61)	8.1	6.97	5.95

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.14: Production of Cotton in Behira Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>MA(1)-F2</u> $w_t = a_t - 0.8979 * a_{t-1}$ (4.99)	8.1	947751 1111351	973990 920920
Housh Aissa	<u>AR(3)-F2</u> $w_t = 0.74 * w_{t-1} - 0.45 * w_{t-2}$ (2.88) (-1.17) $+ 0.71 * w_{t-3}$ (2.47)	9.9	46982	43559
Damanhour	<u>AR (1)-F1</u> $w_t = 0.91 * w_{t-1}$ (3.14)	6.0	163886	155897
Koom Hamada	<u>MA (1)-F1</u> $w_t = a_t + 0.904 * a_{t-1}$ (4.93)	13.6	24648	19362

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.15: Cost of Production in Behira Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>Wheat: AR(1)-F1</u> $w_t = -0.7923 * w_{t-1}$ (5.0)	14.3	747	731
	<u>Rice: AR(1)-F1</u> $w_t = 0.6146 * w_{t-1}$ (2.67)	6.9	965	1080
	<u>Maize: MA (1)-F1</u> $w_t = a_t + 0.86 * a_{t-1}$ (3.34)	6.4	779	1055
	<u>Cotton : MA(2)-F1</u> $w_t = a_t + 0.27a_{t-1} + 0.59*a_{t-2}$ (1.17) (2.56)	6.2	1136	1211

The values within parenthesis are the t-ratios

* Significant at 5%.

6.4 Area, Yield, Production, and Cost of Production in Dakahlia

6.4.1 Wheat

All the time series of the area and yield are nonstationary in the mean. These nonstationary series are transformed by first differencing once or at most two times to get stationary series. All the time series of the production are nonstationary both in the mean and variance. In order to remove the nonstationarity in the production data, we applied the logarithmic transformation followed by differencing the logarithmic series one or two times.

The appropriate identified models for the stationary series are in most cases (nine out of twelve) either AR or MA models of at most third order. The identified models are accepted because they have the lowest chi-square statistics even at 95% significance level.

The forecast error does not exceed

- 1% in most cases of the area models.
- 5% in most cases of the yield models.
- 10% in all cases of the production models.

6.4.2 Rice

Eight out of twelve time series for the area, yield, and production show obvious change in the mean. i.e., non-stationary in the mean. Therefore, we differenced these eight time series data at most twice to achieve stationarity. The other four time series are nonstationary both in the mean and variance. We applied logarithmic transformation followed by first (or second) differences to convert these four series into stationary series. The identified models for the stationary series are promising because they have the lowest chi-square statistics. These appropriate models are (in most cases) either AR or MA models of at most third order.

The forecast error does not exceed

- 12% in all cases of the area models
- 4% in all cases of the yield models
- 9% in all cases of the production models

6.4.3 Maize

All the time series of the yield, and district level area, and district production are nonstationary in the mean. After differencing the yield data once and the area data twice we achieved stationarity. The remaining time series are on the other hand nonstationary in the mean and variance. In order to remove non-stationarity we applied the logarithmic transformation followed by the first differences. The identified models are all promising because they have insignificant chi-square statistics. Nine out of the twelve appropriate models are either AR or MA of at most order 3. The other three models are ARMA models.

The forecast error does not exceed

- 27% in all cases of the area models
- 19% in all cases of the yield models

15% in most cases of the production models

6.4.4 Cotton

Two of the yield time series data are nonstationary in the mean, one series is nonstationary in the variance, and one series is stationary. An unexpected result is that 7 of the 8 time series data for the area and production are found to be stationary. Ten out of the twelve identified models are either AR or MA models of at most order 2. The other two models are ARMA models. These identified models are accepted because the chi-square statistics are insignificant at 95% confidence level.

The forecast error does not exceed

- 16% in all cases of the area models
- 12% in all cases of the yield models
- 14% in all cases of the production models.

6.4.5 The Time Series Models of Total Cost of Production

The total cost time series of Wheat, Rice, Maize, and Cotton are positively trended. i.e., these time series data are nonstationary only in the mean. Therefore, when we differenced the four original series once, it was sufficient to insure stationarity. Cotton required a logarithmic transform to insure stationarity. The four appropriate models for the stationary series are either AR or MA models of at most second order. These models seem most promising because they have the lowest chi-square statistics.

The forecast errors are as follows:

Wheat	4.0%
Rice	0.2%
Maize	6.0%
Cotton	2.0%

Table 6.16: Area of Wheat in Dakahlia Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>M(2)-F1</u> $w_t = a_t - 0.48*a_{t-1} + 0.28a_{t-2}$ (2.36) (1.36)	3.9	213875 218800	207487 216310
El Sinbilaween	<u>AR(2)-F2</u> $w_t = 0.50*w_{t-1} - 0.63*w_{t-2}$ (2.04) (-2.50)	12.3	18258	18272
Sherbin	<u>AR (3)-F2</u> $w_t = - 0.68*w_{t-1} - 0.64w_{t-2}$ (2.89) (-2.61) $+0.61*w_{t-3}$ (2.55)	11.7	19882	19691
El Mansourah	<u>MA(1)-F1</u> $w_t = a_t - 0.48*a_{t-1}$ (2.00)	9.3	22309	22537

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.17: Yield of Wheat in Dakahlia Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>M(3)-F1</u> $w_t = a_t - 0.20a_{t-1} + 0.55a_{t-2}$ (1.09) (0.29) $+ 0.497*a_{t-3}$ (2.53)	8.2	17.05 15.7	17.47 16.62
El Sinbilaween	<u>AR(2)-F1</u> $w_t = 0.89*w_{t-1} - 0.47w_{t-2}$ (-3.78) (-1.99)	8.2	16.08	15.98
Sherbin	<u>AR (2)-F1</u> $w_t = - 0.52*w_{t-1} - 0.18w_{t-2}$ (-2.00) (-0.68)	11.0	15.74	17
El Mansourah	<u>MA(4)-F1</u> $w_t = a_t + 0.05a_{t-1} + 0.27*a_{t-2}$ (0.20) (1.19) $+0.68*a_{t-3} - 0.52*a_{t-4}$ (3.20) (2.00)	7.7	16.04	16.377

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.18: Production of Wheat in Dakahlia Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>AR(3)-LN+F1</u> $w_t = -0.25w_{t-1} + 0.09w_{t-2} + 0.55w_{t-3}$ (-1.36) (0.46) (2.85)	9.4	3646568 3436160	3874782 3269017
El Sinbilaween	<u>ARMA(2,1)-LN+F1</u> $w_t = 0.33w_{t-1} + 0.67w_{t-2} + a_t - 0.91a_{t-1}$ (1.41) (2.88) (3.26)	16.6	293569	299539
Sherbin	<u>ARMA (2,1)-LN+F1</u> $w_t = -0.80w_{t-1} + 0.21w_{t-2} + a_t - 0.99a_{t-1}$ (2.99) (0.64) (4.49)	8.1	313022	327748
El Mansourah	<u>ARMA(2,2)-LN+F1</u> $w_t = -0.85w_{t-1} - 0.65w_{t-2} + a_t + 0.48a_{t-1} + 0.81a_{t-2}$ (-2.16) (-1.86) (1.07) (2.61)	13.2	357872	393328

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.19: Area of Rice in Dakahlia Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>AR(2)-F2</u> $w_t = -0.88 * w_{t-1} - 0.60 * w_{t-2}$ (5.08) (-3.48)	12.1	412198 453796	433721 398917
El Sinbilaween	<u>ARMA(1,1)-F2</u> $w_t = -0.39 w_{t-1} + a_t - 0.97 * a_{t-1}$ (-1.45) (6.97)	3.4	47275	43568
Sherbin	<u>ARMA (3,1)-LN+F1</u> $w_t = -0.66 * w_{t-1} - 0.07 w_{t-2}$ (1.94) (-0.22) $+0.41 w_{t-3} + a_t - 0.85 * a_{t-1}$ (1.45) (2.63)	3.9	39413	35632
El Mansourah	<u>MA(3)-LN+F1</u> $w_t = a_t - 0.096 a_{t-1} - 0.09 a_{t-2}$ (0.38) (0.36) $+ 0.81 * a_{t-3}$ (3.18)	7.7	49377	53691

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.20: Yield of Rice in Dakahlia Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>ARMA(3,2)-F1</u> $w_t = 0.58w_{t-1} + 0.18w_{t-2}$ (-1.85) (0.73) $+0.26w_{t-3} + a_t - 0.26a_{t-1}$ (1.16) (1.16) $+ 0.81*a_{t-2}$ (3.08)	9.5	3.654 3.654	3.404 3.667
El Sinbilaween	<u>AR(1)-LN+F2</u> $w_t = -0.4619*w_{t-1}$ (-1.95)	5.4	3.587	3.717
Sherbin	<u>MA (2)-F1</u> $w_t = a_t + 0.0001a_{t-1} + 0.49*a_{t-2}$ (0.001) (2.10)	9.1	3.862	3.722
El Mansourah	<u>ARMA(1,2)-F1</u> $w_t = -0.70*w_{t-1} + a_t + 0.74*a_{t-1}$ (-1.99) (1.94) $-0.31a_{t-2}$ (0.90)	9.9	3.588	3.694

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.21: Production of Rice in Dakahlia Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>AR(1)-F1</u> $w_t = 0.46 * w_{t-1}$ (2.52)	9.5	1506171 1658171	1410854 1518097
El Sinbilaween	<u>AR(2)-Ln+F1</u> $w_t = 0.94 * w_{t-1} + 0.07 w_{t-2}$ (3.50) (0.24)	3.5	169574	168689
Sherbin	<u>MA(1)-F2</u> $w_t = a_t - 0.7139 * a_{t-1}$ (3.79)	3.5	152213	139299
El Mansourah	<u>MA(3)-LN+F1</u> $w_t = a_t - 0.23 a_{t-1} - 0.27 a_{t-2}$ (0.89) (1.06) $+ 0.77 * a_{t-3}$ (2.81)	8.6	177167	192914

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.22: Area of Maize in Dakahlia Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>ARMA(1.3)-LN+F1</u> $w_t = -0.95w_{t-1} + a_t + 0.36a_{t-1}$ (-2.14) (0.85) $-0.98a_{t-2} - 0.48a_{t-3}$ (2.59) (2.08)	14.3	68255 53420	74239 67914
El Sinbilaween	<u>MA(1)-F2</u> $w_t = a_t + 0.9054a_{t-1}$ (6.11)	10.9	2799	2884
Sherbin	<u>AR (1)-F2</u> $w_t = -0.8019w_{t-1}$ (-4.69)	6.8	6393	6218
El Mansourah	<u>AR(1)-F2</u> $w_t = -0.7120w_{t-1}$ (-3.78)	14.1	4070	4739

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.23: Yield of Maize in Dakahlia Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>ARMA(1,2)-F1</u> $w_t = -0.37w_{t-1} + a_t - 0.44a_{t-1}$ (-1.64) (-3.06) $+ 0.93a_{t-2}$ (6.84)	10.0	21.85 24.83	21.54 20.28
El Sinbilaween	<u>AR(2)-F1</u> $w_t = -0.61w_{t-1} - 0.27w_{t-2}$ (-2.34) (-0.93)	5.1	24.62	21.03
Sherbin	<u>AR (1)-F1</u> $w_t = -0.5012w_{t-1}$ (-2.16)	14.3	24.94	20.03
El Mansourah	<u>MA(3)-F1</u> $w_t = a_t - 0.69a_{t-1} - 0.37a_{t-2}$ (2.44) (1.02) $+ 0.97a_{t-3}$ (3.03)	10.2	25.29	24.47

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.24: Production of Maize in Dakahlia Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>AR(1)-LN+F1</u> $w_t = -0.3746 * w_{t-1}$ (-1.97)	9.5	1622122 1326419	1580892 1589611
El Sinbilaween	<u>MA(3)-LN+F1</u> $w_t = a_t - 0.399 * a_{t-1} - 0.66 * a_{t-2}$ (1.53) (2.57) $+ 0.41 a_{t-3}$ (1.5)	8.0	68925	104297
Sherbin	<u>MA (1)-F1</u> $w_t = a_t - 0.4501 * a_{t-1}$ (1.94)	6.7	157353	155085
El Mansourah	<u>ARMA(2,1)-LN+F1</u> $w_t = 0.44 w_{t-1} + 0.18 w_{t-2} + a_t$ (1.03) (0.52) $- 0.9004 * a_{t-1}$ (2.30)	6.8	102933	118373

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.25: Area of Cotton in Dakahlia Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>AR(1)</u> $w_t = 0.91 * w_{t-1}$ (7.90)	13.8	94062 130271	109503 109349
El Sinbilaween	<u>MA(2)-LN+F1</u> $w_t = a_t - 0.76 * a_{t-1} + 0.75 * a_{t-2}$ (-2.98) (2.76)	7.7	7699	7435
Sherbin	<u>AR (1)</u> $w_t = 0.91 * w_{t-1}$ (4.84)	13.0	14326	14306
El Mansourah	<u>AR(1)</u> $w_t = 0.92 w_{t-1}$ (3.53)	16.4	4614	4605

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.26: Yield of Cotton in Dakahlia Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>AR(1)-F2</u> $w_t = -0.6294 * w_{t-1}$ (-2.5)	13.0	4.89 4.87	5.51 5.08
El Sinbilaween	<u>ARMA(1,2)-LN</u> $w_t = 0.90 * w_{t-1} + a_t + 0.9 * a_{t-1}$ (11.9) (3.98) $+ 0.8 * a_{t-2}$ (3.19)	11.9	4.91	4.45
Sherbin	<u>MA (1)-F2</u> $w_t = a_t - 0.872 * a_{t-1}$ (3.02)	8.5	5.04	4.93
El Mansourah	<u>ARMA(1,1)</u> $w_t = 0.91 * w_{t-1} + a_t - 0.23 a_{t-1}$ (6.12) (0.78)	10.8	5.13	4.6

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.27: Production of Cotton in Dakahlia Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>AR(1)</u> $w_t = 0.92 * w_{t-1}$ (4.39)	10.0	635032	634057
El Sinbilaween	<u>AR(1)</u> $w_t = 0.90 * w_{t-1}$ (5.93)	5.7	37772	37709
Sherbin	<u>AR (2)</u> $w_t = 0.92 * w_{t-1} + 0.08 w_{t-2}$ (3.22) (0.29)	13.1	72132	62186
El Mansourah	<u>AR(1)</u> $w_t = 0.9085 * w_{t-1}$ (4.39)	10.0	635032	634057

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.28: Cost of Production in Dakahlia Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>Wheat: MA(1)-F1</u> $w_t = a_t + 0.875 * a_{t-1}$ (3.53)	7.7	1062	1110
	<u>Rice: MA(1)-F1</u> $w_t = a_t + 0.5844 * a_{t-1}$ (2.36)	11.9	1057	1055
	<u>Maize: MA (1)-F1</u> $w_t = a_t + 0.6199 * a_{t-1}$ (2.79)	11.9	1064	1124
	<u>Cotton: MA(2)-LN+F2</u> $w_t = a_t - 0.3263 a_{t-1}$ (1.28) $- 0.2553 * a_{t-2}$ (2.18)	10.9	1029	1007

The values within parenthesis are the t-ratios

* Significant at 5%.

6.5 Area, Yield, Production, and Cost of Production in Beni Suf

The raw time series data on the governorate level consist of annual data from 1970 to 1997, while the raw data on the district levels consist of annual data from 1980 to 1997. It should be mentioned that rice is not cultivated as a principal crop in Beni Suf. Therefore, the three main crops under study are wheat, maize, and cotton.

The summary results of the 39 estimated models are given in tables (6-29) through (6-38) below.

6.5.1 Wheat

The twelve original time series of the area, yield, and production show obvious change in the mean. i.e., these twelve time series data are nonstationary in the mean. Therefore, we differenced all these raw time series data at most twice to achieve stationarity. The appropriate identified models for the stationary series are in most cases (ten out of twelve) either AR or MA models of at most second order. The remaining two models are ARMA(1,1) model. The appropriate models are all promising because they have the lowest chi-square statistics.

The forecast error does not exceed

- 5% in all cases of the area models.
- 6% in most cases of the yield models.
- 11% in all cases of the production models.

6.5.2 Maize

About half of the original time series of the area, yield, and production show nonstationarity in the mean. While the other half indicate nonstationarity both in the mean and variance. Under the first condition, we differenced the raw time series data at most twice to achieve stationarity. While under the second condition we applied logarithmic transformations followed by second differences to achieve stationarity. All identified models (except three) are either AR(1) or MA(1). The other three models are ARMA models. The appropriate identified models for the stationary series are accepted because the chi-square statistics are insignificant at 95% confidence level.

The forecast error does not exceed

- 23% in all cases if the area models.
- 17% in all cases of the yield models
- 21% in all cases of the production models.

6.5.3 Cotton

All original time series of the area, yield, and production are nonstationary in the mean only. Therefore, we differenced these original time series data once or twice to achieve stationarity. Ten out of twelve appropriate models for the stationary series are either AR or MA models of

at most order 2. The other two appropriate models are ARMA models. All of these models are promising because they have the lowest chi-square statistics.

The forecast error does not exceed

- 11% in most cases of the area models
- 14% in all cases of the yield models
- 22% in all cases of the production models

6.5.4 The Time Series Models of Total Cost of Production

The situation here is as similar as in Behira and Dakahlia. The total cost time series of wheat, maize, and cotton are positively trended. i.e., these raw time series data are nonstationary in the mean only. Therefore, when we differenced the three original series twice, we found it sufficient to ensure stationarity. The three identified models for the stationary series are MA(1) models. These models seem very promising because they have the lowest chi-square statistics.

The forecast error are as follows:

- Wheat 8%
- Maize 6%
- Cotton 3%

Table 6.29: Area of Wheat in Beni Suf Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>AR(1)-F2</u> $w_t = -0.6767 * w_{t-1}$ (-4.28)	11.2	104488 114350	107472 110648
Ahnasia	<u>MA(1)-F2</u> $w_t = a_t - 0.9106 * a_{t-1}$ (-4.28)	8.4	22213	21919
Beni Suef	<u>MA(1)-F2</u> $w_t = a_t - 0.9168 * a_{t-1}$ (-4.85)	2.4	13693	13239
El Fashn	<u>MA(1)-F2</u> $w_t = a_t - 0.9482 * a_{t-1}$ (5.76)	6.5	17246	16357

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.30: Yield of Wheat in Beni Suef Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>MA(1)-F2</u> $w_t = a_t - 0.95*a_{t-1}$ (9.24)	9.0	15.86 16.72	15.48 16.46
Ahnasia	<u>AR(1)-F2</u> $w_t = -0.5466*w_{t-1}$ (-2.44)	11.4	15.73	15.63
Beni Suef	<u>AR (2)-F2</u> $w_t = -0.72*w_{t-1} - 0.60*w_{t-2}$ (-3.24) (2.7)	5.6	15.37	17.89
El Fashn	<u>MA(1)-F2</u> $w_t = a_t - 0.9338*a_{t-1}$ (5.76)	8.5	16.91	15.84

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.31: Production of Wheat in Beni Suef Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>ARMA(1,1)-F2</u> $w_t = -0.35w_{t-1} + a_t - 0.929*a_{t-1}$ (-1.67) (8.29)	14.1	1657180 1911906	1637037 1741877
Ahnasia	<u>ARMA(1,1)-F2</u> $w_t = -0.48w_{t-1} + a_t - 0.909*a_{t-1}$ (-1.87) (4.24)	9.4	349368	349485
Beni Suef	<u>AR (1)-F1</u> $w_t = -0.5121*w_{t-1}$ (-2.31)	2.8	210483	214380
El Fashn	<u>MA(1)-F2</u> $w_t = a_t - 0.906*a_{t-1}$ (-4.45)	6.6	291572	258439

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.32: Area of Maize in Beni Suef Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>AR(1)-F2</u> $w_t = -0.7277 * w_{t-1}$ (-4.82)	17.2	100850	106371
Ahnasia	<u>MA(1)-LN+F2</u> $w_t = a_t - 0.9078 * a_{t-1}$ (-6.13)	6.4	16197	12419
El Fashn	<u>MA (1)-F2</u> $w_t = a_t - 0.9069 * a_{t-1}$ (-5.68)	11.7	25859	29270

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.33: Yield of Maize in Beni Suef Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>ARMA(4,1)-F2</u> $w_t = -0.64 * w_{t-1} - 0.54 * w_{t-2}$ (-2.60) (-2.23) $-0.63 * w_{t-3} - 0.26 w_{t-4}$ (-2.62) (-1.12) $+ a_t - 0.9109 * a_{t-1}$ (8.40)	10.4	18.75 20.28	16.71 21.346
Ahnasia	<u>MA(1)-LN+F2</u> $w_t = a_t - 0.9399 * a_{t-1}$ (-5.55)	10.5	22.43	18.58
El Fashn	<u>MA (1)-F2</u> $w_t = a_t - 0.879 * a_{t-1}$ (3.48)	13.6	18.71	18.63

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.34: Production of Maize in Beni Suef Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>ARMA(1,3)-LN+F2</u> $w_t = -0.80w_{t-1} + a_t - 0.85a_{t-1}$ (-4.04) (-4.32) $-0.92a_{t-2} + 0.79a_{t-3}$ (-3.99) (5.10)	15.2	1948181 2045472	1533403 2594748
Ahnasia	<u>MA(1)-LN+F2</u> $w_t = a_t - 0.9052a_{t-1}$ (6.21)	12.7	363330	233281
El Fashn	<u>ARMA (1,1)-F2</u> $w_t = -0.66w_{t-1} + a_t - 0.95a_{t-1}$ (-2.45) (-4.65)	16.3	483748	457257

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.35: Area of Cotton in Beni Suef Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>ARMA(3,1)-F1</u> $w_t = 0.30w_{t-1} + 0.24w_{t-2}$ (1.30) (1.11) $+0.45w_{t-3} + a_t - 0.86a_{t-1}$ (1.98) (6.34)	5.2	44764 54594	45803 54671
Ahnasia	<u>MA(1)-F2</u> $w_t = a_t - 0.9033a_{t-1}$ (5.57)	9.8	13477	9472
Beni Suef	<u>AR (2)-F2</u> $w_t = -0.36w_{t-1} - 0.64w_{t-2}$ (-1.55) (-2.66)	13.1	7838	7125
El Fashn	<u>AR(2)-F1</u> $w_t = -0.65w_{t-1} - 0.37w_{t-2}$ (-2.42) (-1.20)	3.3	9776	8711

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.36: Yield of Cotton in Beni Suef Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>AR(2)-F2</u> $w_t = -0.84 * w_{t-1} - 0.52 w_{t-2}$ (-2.69) (-1.4)	8.9	7.25	6.41
Ahnasia	<u>MA(1)-F2</u> $w_t = a_t - 0.9268 * a_{t-1}$ (-3.96)	13.4	6.46	6.054
Beni Suef	<u>MA (1)-F1</u> $w_t = a_t - 0.4907 * a_{t-1}$ (2.03)	9.7	7.6	6.52
El Fashn	<u>ARMA(2,1)-F2</u> $w_t = -0.41 w_{t-1} - 0.65 * w_{t-2} + a_t$ (-1.27) (-1.94) $- 0.93 * a_{t-1}$ (4.93)	5.9	7.26	6.27

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.37: Production of Cotton in Beni Suef Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>AR(1)-F2</u> $w_t = 0.91 * w_{t-1}$ (6.62)	18.2	251602 396031	308850 308382
Ahnasia	<u>AR(2)-F1</u> $w_t = -0.779 * w_{t-1} - 0.46 w_{t-2}$ (-2.96) (-1.69)	5.5	87120	71313
Beni Suef	<u>AR (2)-F2</u> $w_t = - 0.96 * w_{t-1} - 0.669 * w_{t-2}$ (-3.67) (-2.21)	7.8	59564	53240
El Fashn	<u>AR(2)-F1</u> $w_t = - 0.67 * w_{t-1} - 0.71 * w_{t-2}$ (-2.46) (-2.24)	4.4	71003	65913

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.38: Cost of Production in Beni Suef Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>Wheat: MA(1)-F2</u> $w_t = a_t - 0.5558 * a_{t-1}$ (-2.52)	9.9	1003	921
	<u>Maize: MA(1)-F1</u> $w_t = a_t + 0.6199 * a_{t-1}$ (2.79)	11.9	1064	1124
	<u>Cotton: MA(1)-F2</u> $w_t = a_t - 0.9115 * a_{t-1}$ (4.40)	5.4	1090	1054

The values within parenthesis are the t-ratios

* Significant at 5%.

6.6 Area, Yield, Production, and Cost of Production in Assiut Governorate

The original time series data on the governorate level consist of annual data from 1970 to 1997, while the original data on the district levels consist of annual data from 1980 to 1997.

As in Beni Suef governorate rice is not cultivated in Assiut governorate. Hence, the three main crops under study are wheat, maize, and cotton.

The summary results of the 39 estimated models are given in tables (6-39) through (6-48).

6.6.1 Wheat

Almost all original time series (eleven out twelve) of the area, yield, and production show obvious change in the mean .i.e., these time series data are nonstationary in the mean. Therefore, we differenced all these raw time series data once or twice to achieve stationarity. The appropriate identified models for the stationary series are in most cases (eleven out of twelve) either AR or MA models of at most second order. The remaining model is ARMA(1,2) model. The appropriate models are all promising because they have the lowest chi-square statistics.

The forecast error does not exceed

- 13% in all cases of the area models.
- 7% in all cases of the yield models.
- 9% in all cases of the production models.

6.6.2 Maize

The original time series of the yield, and half of the area and production are nonstationary in the mean only and the series of the second half of the area and production are nonstationary in both of the mean and the variance. The nonstationary series in the mean are converted to stationary series by differencing the original series once or twice. The nonstationarity in the variance also needs, logarithmic transformation. The appropriate identified models consist of a group of ARMA (p,q) models. Where p and q are at most of order 2. All of these models are promising because they have the lowest chi-square statistics even at 95% level.

The forecast error does not exceed

- 8% in most cases if the area models.
- 11% in all cases of the yield models
- 7% in all cases of the production models.

6.6.3 Cotton

Of the original time series (10 of 12) of the area, yield, and production are nonstationary in the mean only. Therefore, we differenced these original time series data at most twice to achieve stationarity. The appropriate models for the stationary series are either AR or MA models of at most order 2. These models are all accepted because they have insignificant chi-square statistics at 95% level.

The forecast error does not exceed

29% in most cases of the area models

18% in all cases of the yield models

22% in all cases of the production models

6.6.4 The Time Series Models of Total Cost of Production

The main findings here are as similar as in the other three governorates. The total cost time series of wheat, maize, and cotton are positively trended. i.e., these original time series data are nonstationary in the mean only. Differencing the three original series twice at most, results in stationary time series. The three identified models for the stationary series are pure MA or AR of at most order two. These models seem very promising because they have the lowest chi-square statistics.

The forecast errors are as follows:

Wheat 1.0%

Maize 12.0%

Cotton 7.0%

Table 6.39: Area of Wheat in Assiut Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>MA(1)-F2</u> $w_t = a_t - 0.926 * a_{t-1}$ (-5.05)	15.3	133062 130999	139353 129564
Abnab	<u>AR(1)-F2</u> $w_t = -0.6878 * w_{t-1}$ (-3.13)	5.6	17163	17166
Abou-Tieg	<u>MA (1)-F2</u> $w_t = a_t - 0.9003 * a_{t-1}$ (-3.57)	7.5	15635	13558
Assuit	<u>MA(2)-LN+F2</u> $w_t = a_t - 0.95 * a_{t-1} - 0.008 a_{t-2}$ (-2.85) (0.02)	15.1	13709	17676

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.40: Yield of Wheat in Assiut Governorate

Governorate	Estimated Model	Chi-square	The Last Year
-------------	-----------------	------------	---------------

level & Districts		Statistic	Actual Value	Forecast Value
Total Governorate	<u>MA(1)-F2</u> $w_t = a_t - 0.9277 * a_{t-1}$ (7.27)	9.6	16.49 17.87	16.6 16.67
Abnob	<u>MA(2)-F1</u> $w_t = a_t + 0.60 * a_{t-1} - 0.34 a_{t-2}$ (2.16) (1.24)	6.3	17.52	16.42
Abou-Tieg	<u>ARMA (1,2)-F1</u> $w_t = -0.86 * w_{t-1} + a_t$ (-2.79) $+ 0.72 * a_{t-1} - 0.51 a_{t-2}$ (2.70) (1.45)	7.5	16.22	15.77
Assuit	<u>MA(2)-F1</u> $w_t = a_t - 0.80 * a_{t-1} + 0.59 * a_{t-2}$ (5.25) (2.23)	3.5	18.65	16.72

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.41: Production of Wheat in Assiut Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>AR(1)-F2</u> $w_t = -0.6543 * w_{t-1}$ (-2.68)	7.5	2194192 2340962	2220945 2523208
Abnoub	<u>AR(1)-F2</u> $w_t = 0.7137 * w_{t-1}$ (-3.37)	6.5	300650	282362
Abou-Tieg	<u>MA (3)-F1</u> $w_t = a_t + 0.78 * a_{t-1} - 0.64 * a_{t-2}$ (2.59) (2.01)	11.0	253623	261258
Assuit	$-0.85 a_{t-3}$ (-2.81) <u>AR(2)-F1</u> $w_t = -0.103 w_{t-1} - 0.546 * w_{t-2}$ (0.40) (-1.98)	6.4	255631	279777

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.42: Area of Maize in Assiut Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>MA(1)-F2</u> $w_t = a_t - 0.9669*a_{t-1}$ (-8.96)	14.8	76497 74654	82735 72886
Abnobe	<u>MA(1)-LN+F2</u> $w_t = a_t - 0.9725*a_{t-1}$ (-7.13)	10.9	2780	3593
Abou-Tieg	<u>ARMA (2,1)-LN+F2</u> $w_t = -0.64*w_{t-1} - 0.88*w_{t-2}$ (-2.98) (-4.57) + $a_t - 0.91*a_{t-1}$ (-4.17)	12.0	4961	5366
Assuit	<u>MA(1)-F2</u> $w_t = a_t - 0.8785*a_{t-1}$ (-3.93)	10.8	17400	17842

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.43: Yield of Maize in Assiut Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>AR(2)-F2</u> $w_t = -0.88*w_{t-1} - 0.57*w_{t-2}$ (-4.29) (-2.52)	11.8	20.12 22.46	19.45 21.37
Abnobe	<u>ARMA(1,1)-F2</u> $w_t = -0.45*w_{t-1} + a_t - 0.91*a_{t-1}$ (-1.73) (4.55)	8.6	22.42	20.38
Abou-Tieg	<u>ARMA (1,1)-F2</u> $w_t = -0.63*w_{t-1} + a_t$ (-2.87) - $0.89*a_{t-1}$ (-4.70)	11.6	20.64	18.71
Assuit	<u>MA(1)-F2</u> $w_t = a_t - 0.8740*a_{t-1}$ (-3.51)	7.5	23.23	20.66

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.44: Production of Maize in Assiut Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>MA(1)-LN+F2</u> $w_t = a_t - 0.9145 * a_{t-1}$ (-6.30)	7.6	1676686	1613635
Abnob	<u>ARMA(1,3)-LN+F2</u> $w_t = -0.17w_{t-1} + a_t - 0.91 * a_{t-1}$ (-0.34) (-1.95) $-0.85a_{t-2} + 0.84 * a_{t-3}$ (1.66) (2.29)	8.7	58513	62330
Abou-Tieg	<u>ARMA (2,1)-F2</u> $w_t = -0.77 * w_{t-1} - 0.75 * w_{t-2} + a_t$ (-3.40) (-3.49) $- 0.89 * a_{t-1}$ (-3.66)	10.2	102387	109126
Assuit	<u>MA(1)-LN+F2</u> $w_t = a_t - 0.8965 * a_{t-1}$ (-4.32)	9.5	404236	375871

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.45: Area of Cotton in Assiut Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>AR(1)-F3</u> $w_t = -0.8184 * w_{t-1}$ (-6.68)	12.9	43748	30802
Abnobe	<u>AR(2)-F2</u> $w_t = -0.71 * w_{t-1} - 0.58 * w_{t-2}$ (-2.75) (-2.12)	9.4	6210	5214
Abou-Tieg	<u>AR(2)-F3</u> $w_t = -0.97 * w_{t-1} - 0.49 w_{t-2}$ (-3.02) (-1.43)	4.7	4015	1993
Assuit	<u>MA(1)-LN+F1</u> $w_t = a_t + 0.8331 * a_{t-1}$ (4.29)	8.6	6875	4915

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.46: Yield of Cotton in Assiut Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>MA(1)-F2</u> $w_t = a_t - 0.9149 * a_{t-1}$ (-8.15)	10.0	7.94 9.44	8.41 7.75
Abnobe	<u>MA(1)-F2</u> $w_t = a_t - 0.9556 * a_{t-1}$ (-6.68)	3.3	8.36	7.51
Abou-Tieg	<u>MA (1)-F2</u> $w_t = a_t - 0.9563 * a_{t-1}$ (-6.74)	8.2	8.34	6.58
Assuit	<u>MA(1)-F2</u> $w_t = a_t - 0.8307 * a_{t-1}$ (-3.58)	5.4	10.41	10.11

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.47: Production of Cotton in Assiut Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>MA(1)-F2</u> $w_t = a_t - 0.908 * a_{t-1}$ (-10.9)	11.6	412964	319328
Abnoub	<u>MA(1)-F3</u> $w_t = a_t - 0.9333 * a_{t-1}$ (-4.20)	11.9	51938	46133
Abou-Tieg	<u>MA (1)-F3</u> $w_t = a_t - 0.9601 * a_{t-1}$ (-5.12)	10.6	33466	23408
Assuit	<u>AR(1)-LN+F3</u> $w_t = - 0.5694 * w_{t-1}$ (-2.40)	10.3	71556	60054

The values within parenthesis are the t-ratios

* Significant at 5%.

Table 6.48: Cost of Production in Assiut Governorate

Governorate level & Districts	Estimated Model	Chi-square Statistic	The Last Year	
			Actual Value	Forecast Value
Total Governorate	<u>Wheat: MA(1)-F1</u> $w_t = a_t + 0.8864 * a_{t-1}$ (6.70)	8.1	755	761
	<u>Maize: MA(2)-F2</u> $w_t = a_t - 0.895 * a_{t-1} + 0.91 * a_{t-2}$ (-3.98) (3.44)	3.3	1205	1055
	<u>Cotton: AR (1)-F2</u> $w_t = - 0.6746 * w_{t-1}$ (-2.02)	16.5	1260	1347

The values within parenthesis are the t-ratios

* Significant at 5%.

6.6 General Results of Time Series Analyses for Data Quality and Accuracy

As it has been mentioned earlier the (TS) analysis will reject the null hypothesis, leading one to conclude that there may be interference in the collection and/or processing of the data, if many or all the variables show the same basic patterns, as measured by the parameters of (TS) analysis. In addition, one would not expect forecasts of the variables to have extremely low levels of error. If these are found, this would be an additional corroboration of unusual consistency in the data.

Based on (TS) analyses of the published agricultural data, one can make the following general remarks:

Plots of the original time series data of the random variables under study (area, yield, and production) for the main field crops (wheat, rice, maize, and cotton) all reflected nonstationarity, as one would expect. However, the unexpected result is that most of these plots look very much the same and indicate the same kind of nonstationarity. That is, they have nonstationarity in the mean and they all exhibit a gradual movement. According to our own knowledge of agricultural, one would expect that the different crops and districts would have different types of nonstationarity and movements. The correlograms of the Autocorrelation Functions (ACF's) for most of the original time series data decay exponentially (slow decrease) very slowly, confirming first unexpected result about the same kind of nonstationarity in most time series under study [see Figure (6-8) below] .

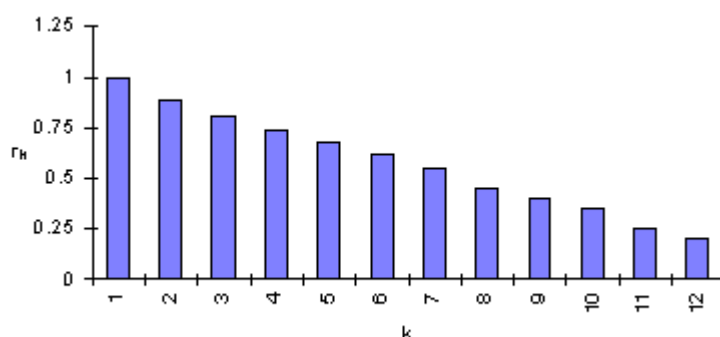


Chart 6.8:

Sample ACF of most original time series under study

The correlograms of the Partial Autocorrelation Functions (PACF's) for most of the original nonstationary data have only a significant single spike at lag one [see Figure (6-9) below].

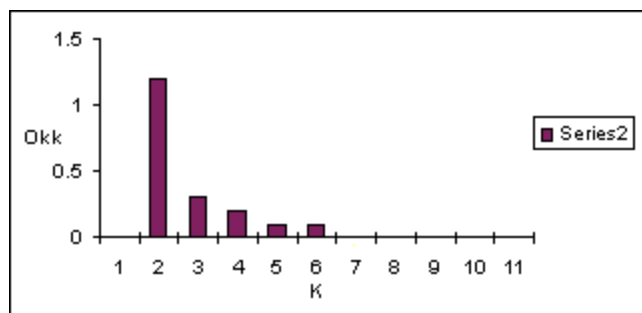


Chart 6.9: Sample PACF of most original time series under study

This suggests the possibility of a first-order autoregressive AR(1) model:

$$X_t = f X_{t-1} + u_t$$

for all 182 original time series data sets under study which is the second unexpected result.

It is well known that X_t will be a nonstationary function if the parameter f (in the above model) equals or exceeds unity. The estimated AR(1) models for the 182 nonstationary series show that $f \sim 1$ in most cases which is the third unexpected result.

Remember that the data for different districts and crops are expected to vary a great deal due to weather, soil, cultural practices, prices and other factors. Therefore, one would expect that different crops and districts would have different types of nonstationarity and movements. When nonstationarity and models are very similar, it indicates that they might not be independent or that there is some influence on the data which does not permit the true variability to take place.

It is interesting to note that the three unexpected results stated in above were more obvious and frequent in the original time series of wheat, maize, and rice, than in the case of cotton.

The forecast errors in many cases are less than 1% which is the fourth unexpected result. Practical experience with time series models based on economic data reveals that such a consistent very low level of forecast errors is very unusual. Again, it should be mentioned that the fourth unexpected result concerning the low levels of the forecast errors is related to the cases of wheat, rice, and maize more than in case of cotton.

In view of the above unexpected result we can, with high probability, state that wheat, rice, and maize time series data are too uniform and thus very likely to be less accurate than cotton data. This result, in our opinion, is due to some interference in collecting and reporting the data at different levels (cooperatives districts-governorates).

The total cost of production is obviously affected by:

- The prices and quantities of the inputs used in the production process,
- The level of output produced,
- The state of technical knowledge used in production.

The directions and variations in the above three factors are not the same. Therefore, it was unexpected result to find that the time series data of total cost of production for each crop lay on a straight line. This result is found to be true for total cost of production data of wheat, rice, maize, and cotton.

Moreover, the total cost of production for all of the four field crops have, in addition to the above results , the following:

- The same type of nonstationarity in all of the total cost of production time series data,
- The same AR(1) model,
- The low levels of forecast errors.

Therefore, based on the above result one can conclude that the total cost of production time series data are too uniform and thus likely not to be accurate. Again, there is some interference during the various steps of collecting, tabulating, and reporting of these data.

7. FINDINGS: STATISTICAL ANALYSES

In theory area, yield and production data, farm gate prices, cost of production and other farmer related data should be gathered at the farm level and aggregated to district, governorate and national level. If data is correct at the farm level there should be no changes made. If an occasional error is found there might be justification to correct it, but only after verification of errors.

An important part of assessing data quality is to determine the level at which the data is most likely to be correct. If there are levels where the correctness of the data is in question, then one should determine the reasons for inaccuracies, if possible. Are there mistakes during transmission, transcription or other processing problems? Are there other influences causing manipulation of the data to meet external or internal goals? To answer these questions, we followed the data through every level as it traveled from the village through district and governorate offices to national publication. Annex A contains all of the detailed data.

7.1 Graphical Investigation of Data

Looking at charts and graphs of data can help one understand the relationships between variables, identify trends not readily apparent and unusual relationships that may warrant further verification or checks. The following graphs are a few of the many that were developed. They are given here with notes to draw attention to interesting considerations.

Chart 7.1: Cotton Area in Behira, Damenhour District-District, Governorate and Published

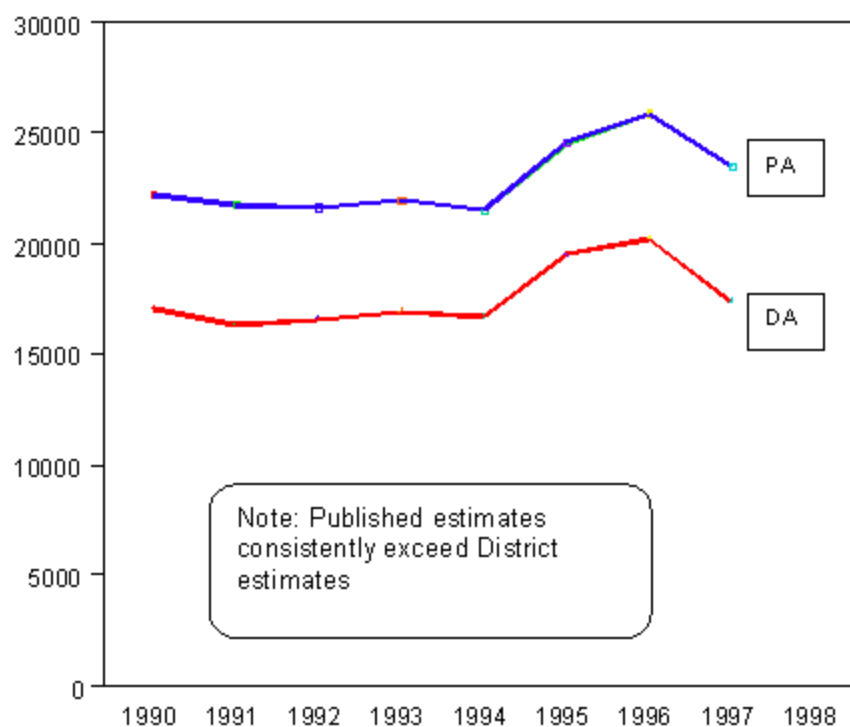


Chart 7.2: Cotton Production in Behira, Damenhour District-District, Governorate and Published

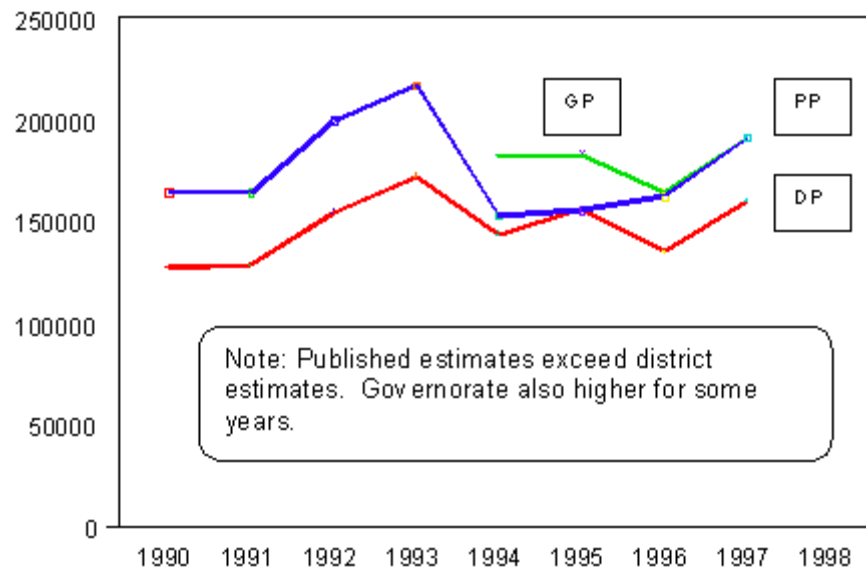


Chart 7.3: Cotton Area in Dakahlia, Mansoura District- District, Governorate and Published

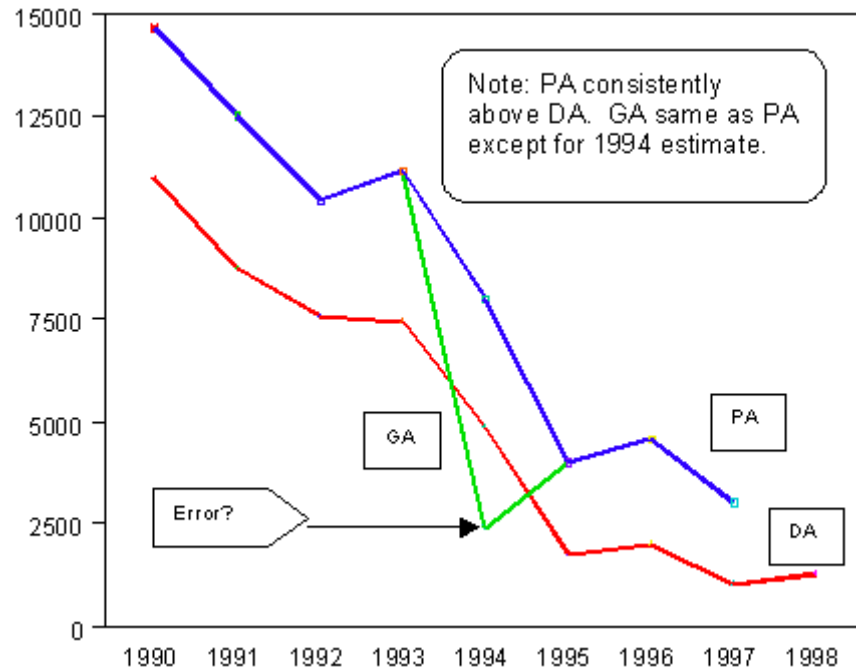


Chart 7.4: Cotton Production in Dakahlia, Mansoura District-District, Governorate, and Published

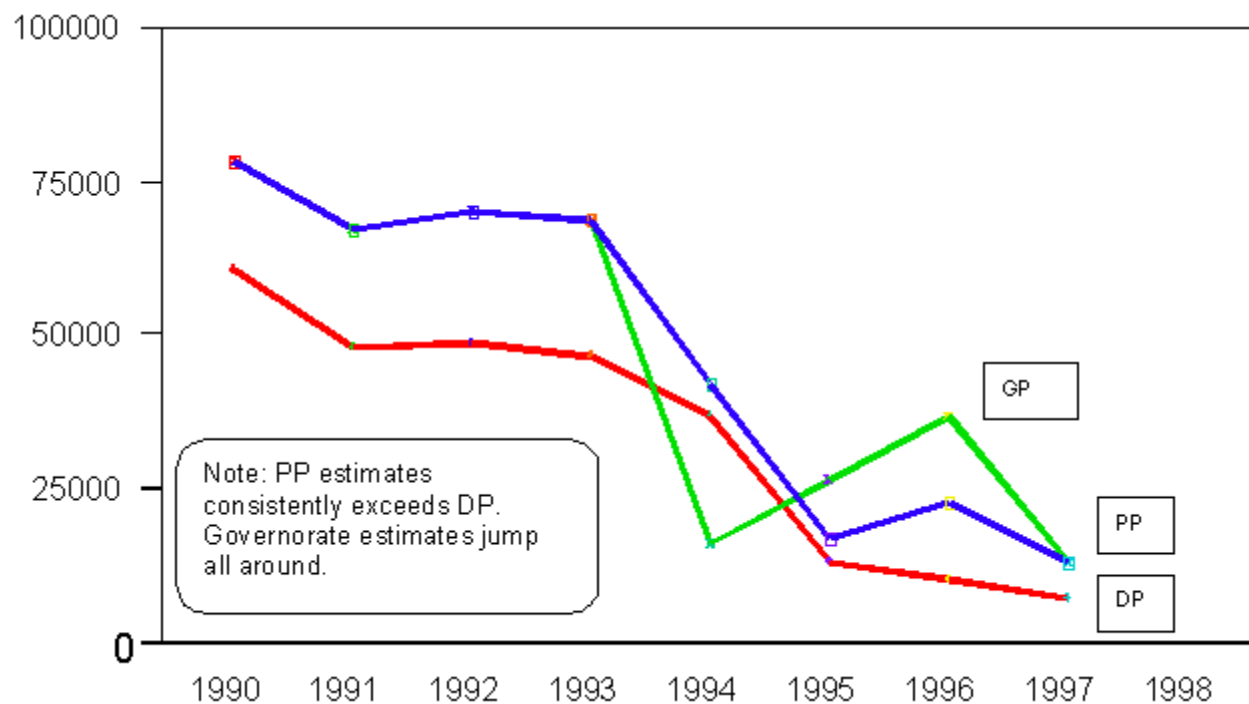


Chart 7.5: Maize in Behira, Damanhour District-District Area/100, Yield, Production/1000

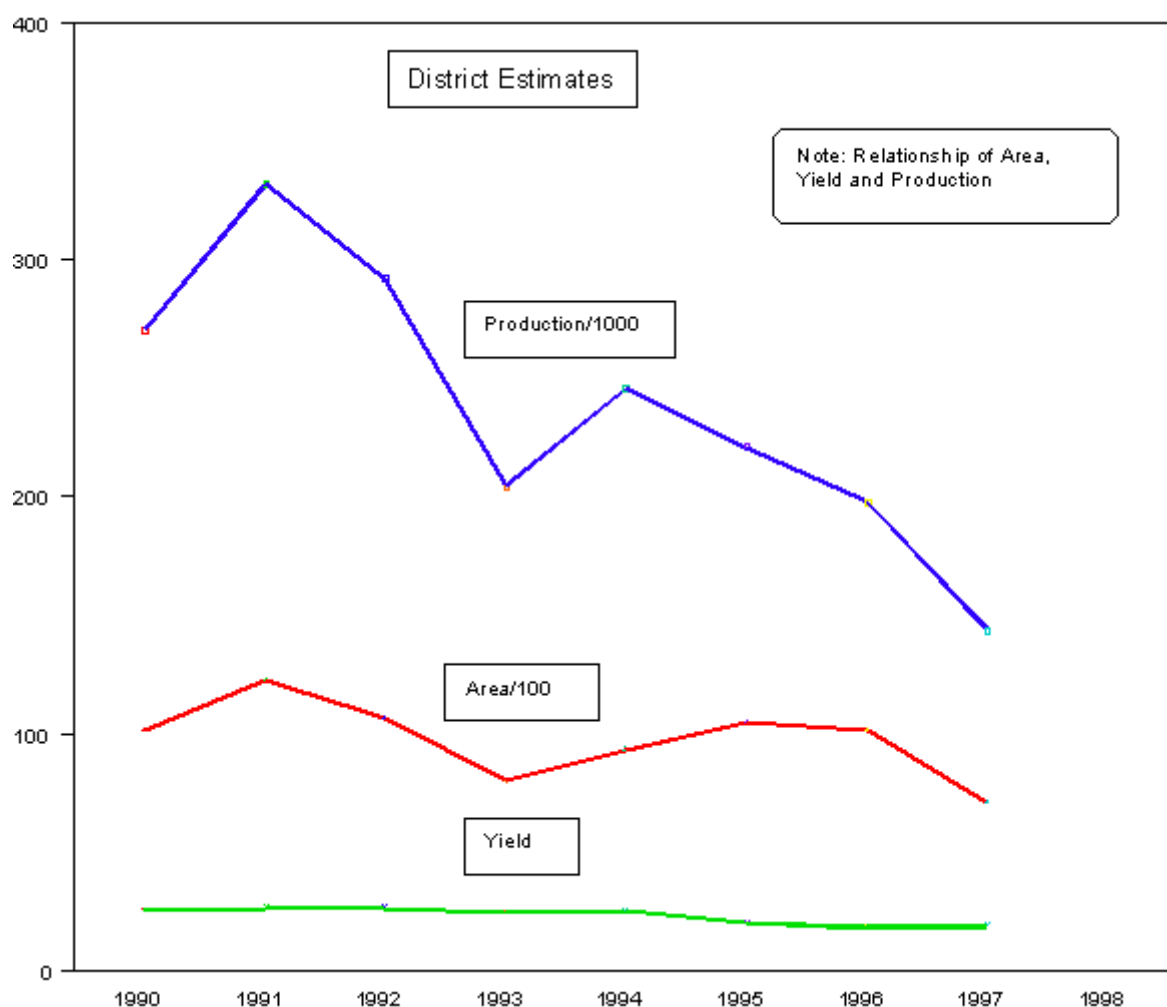


Chart 7.6: Maize in Behira, Damanhour District-Area/100, Yield, Production/1000

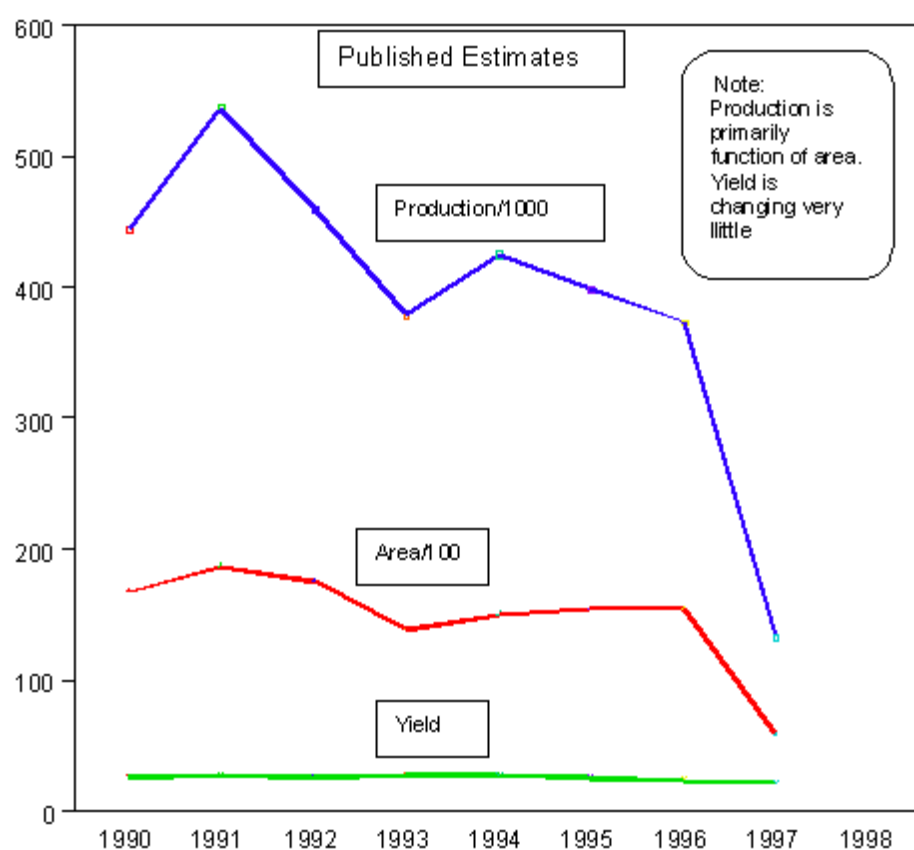


Chart 7.7: Maize in Dakahlia, Sherbin District-Area/100, Yield, Production/1000

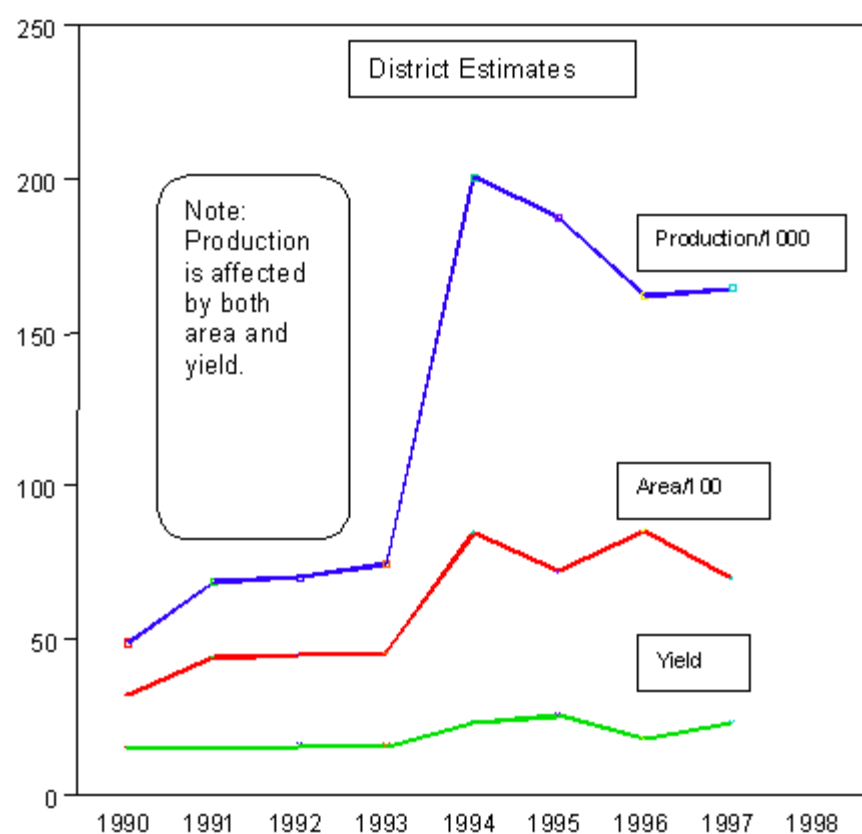


Chart 7.8: Maize in Dakahlia, Sherbin-Area/100, Yield, Production/1000

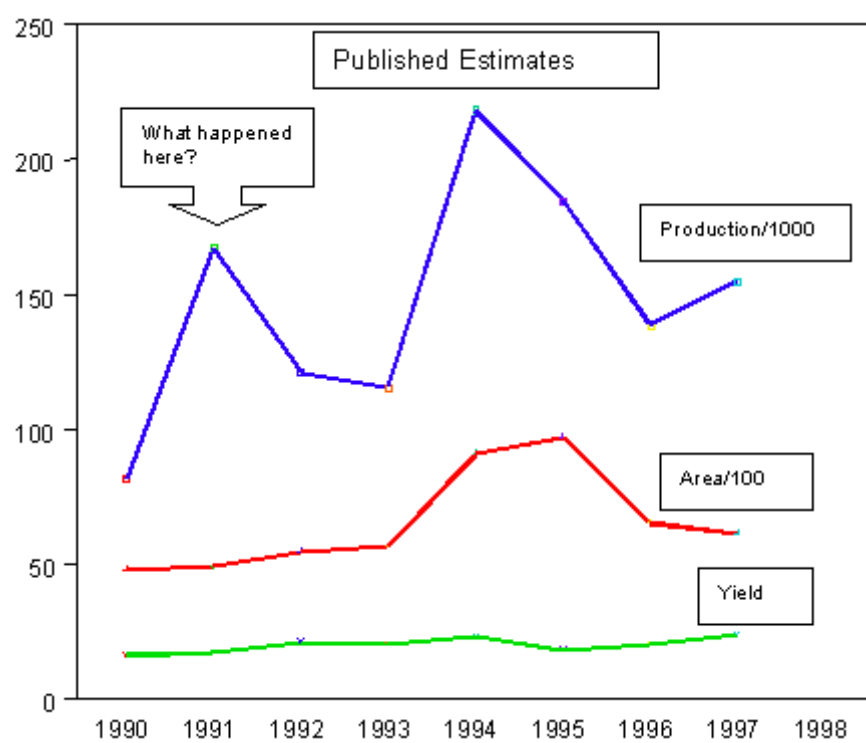


Chart 7.9: Rice Area in Behira, Kom Hamada District-District, Governorate and Published Estimates

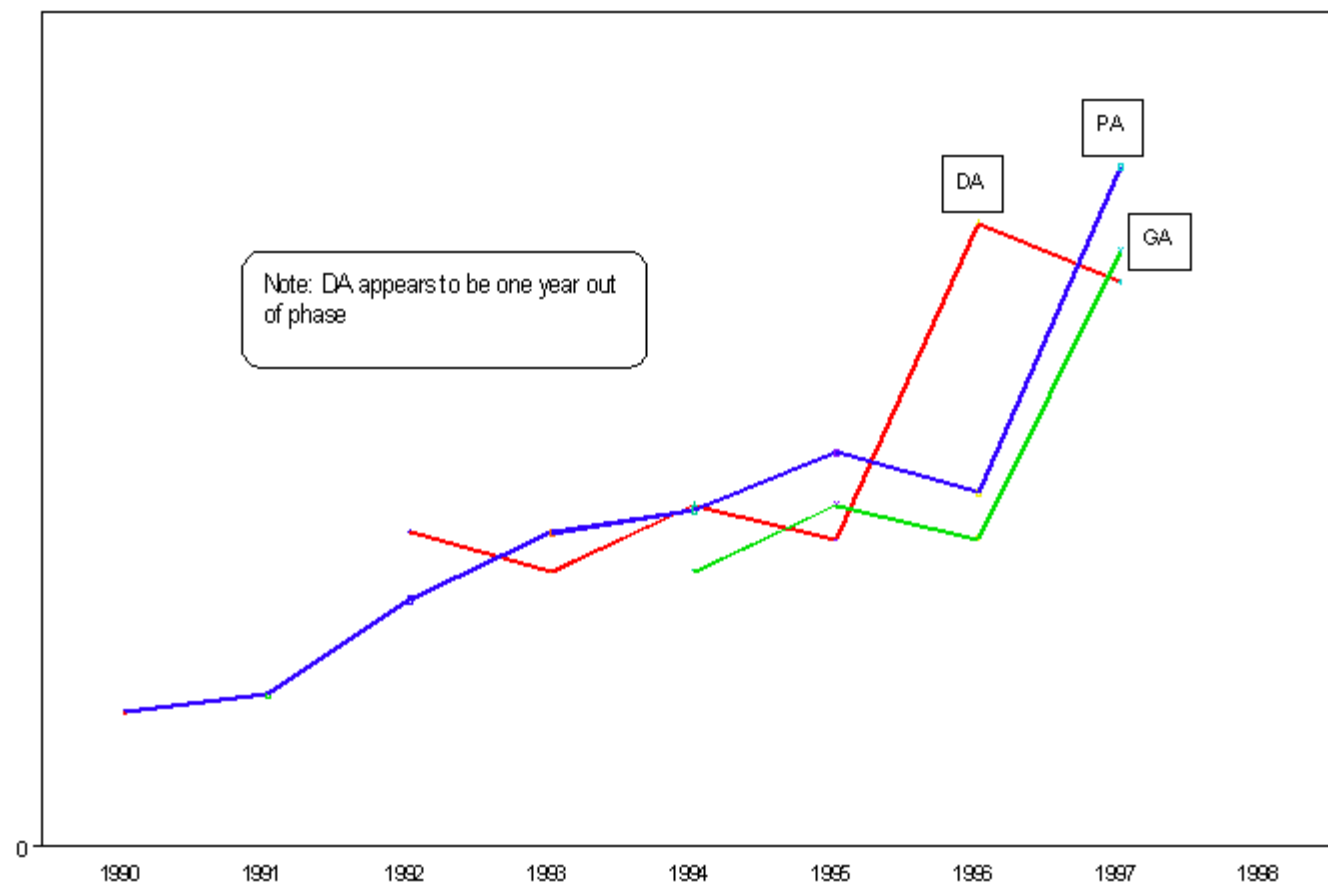


Chart 7.10: Rice Production in Behira, Kom Hamada District-District, Governorate and Published Estimates

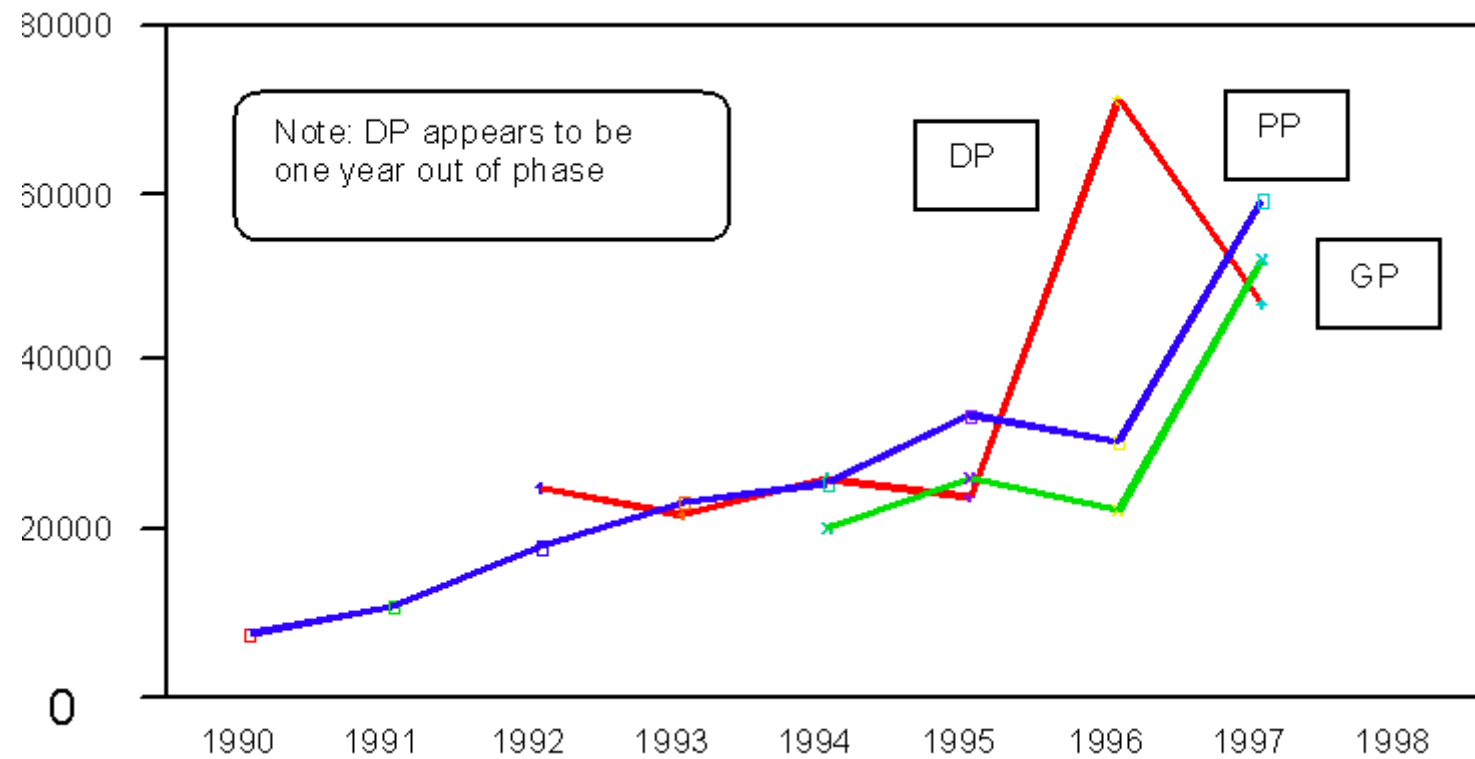


Chart 7.11: Rice Area in Behira, Damanhour District-District, Governorate and Published Estimates

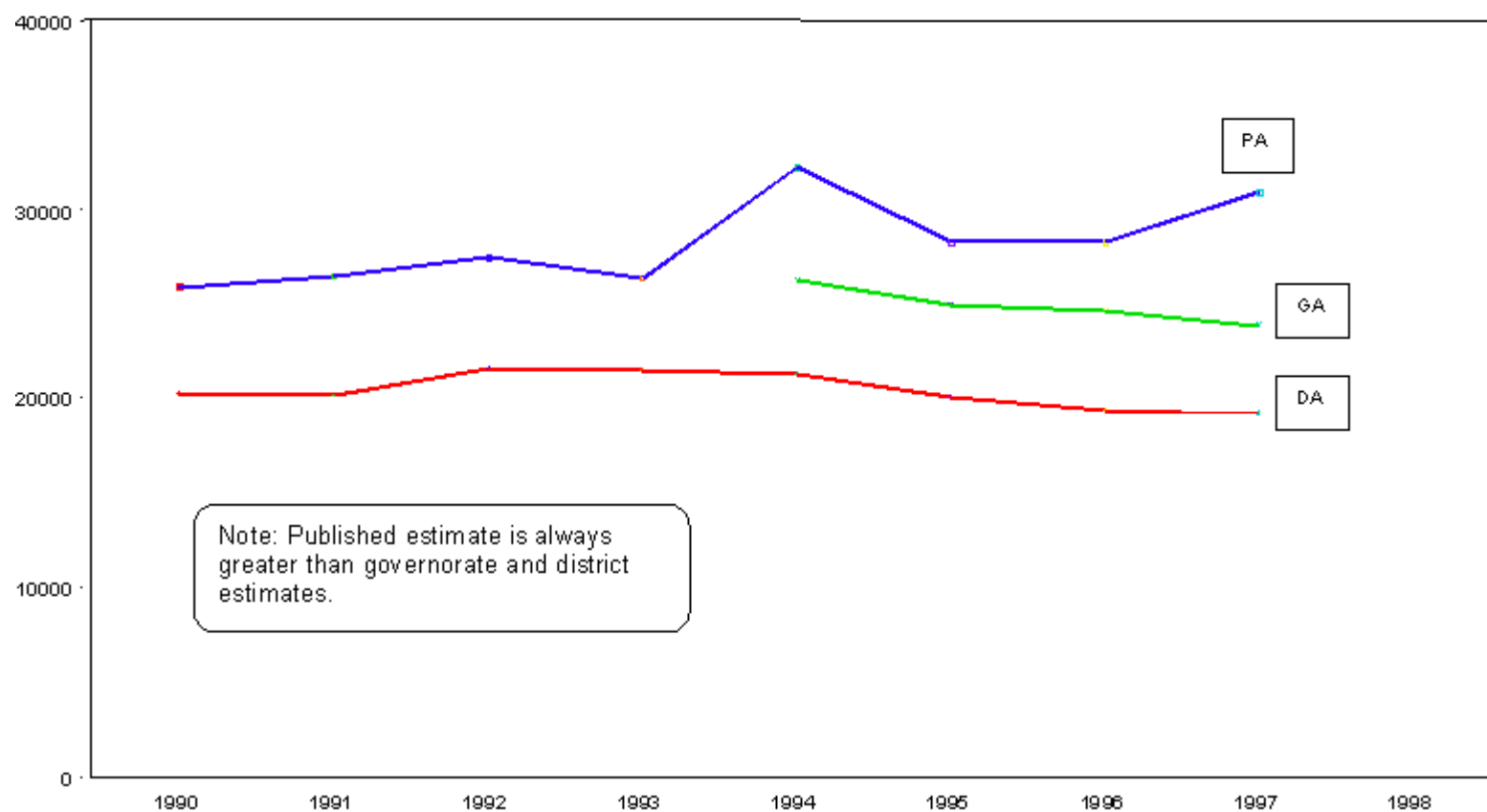


Chart 7.12: Rice Production in Behria, Damanhour District-District, Governorate and Published Estimates

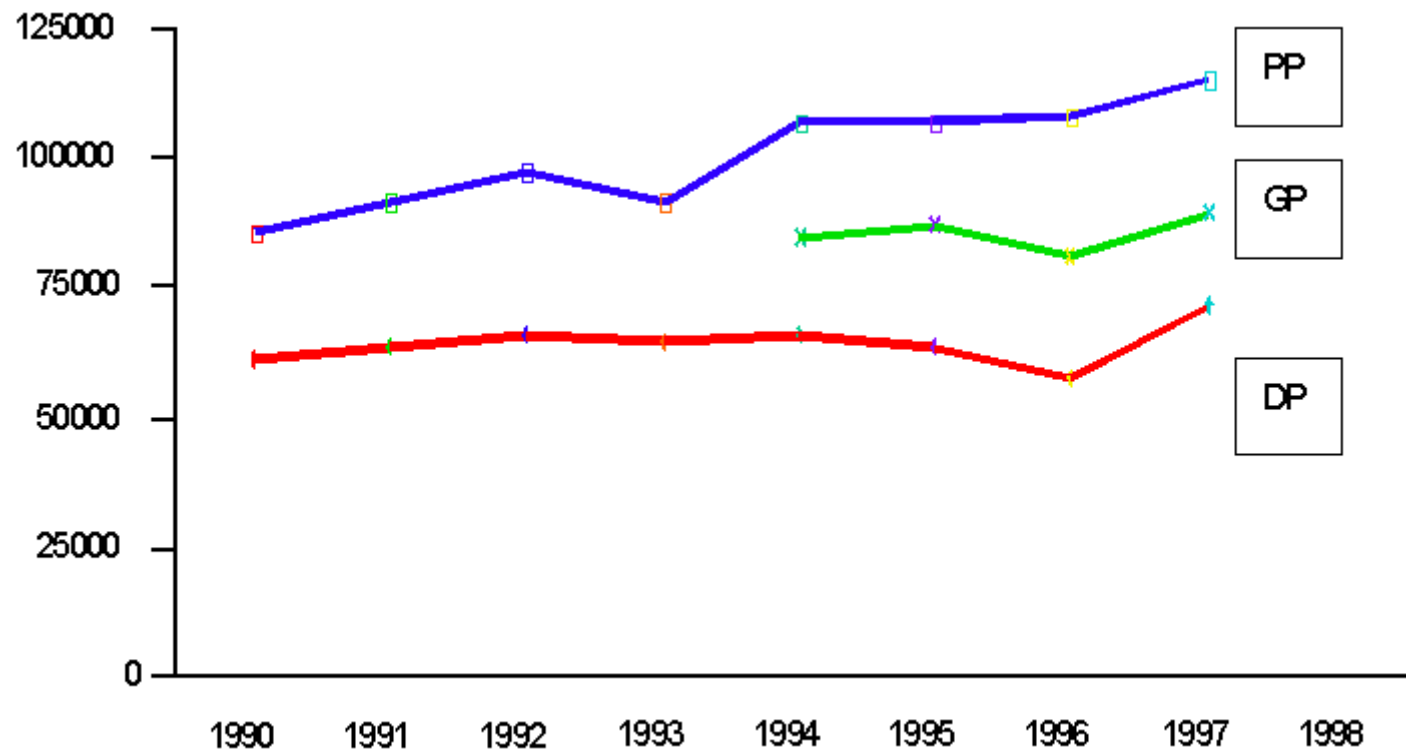


Chart 7.13: Rice Area in Dakalahia, Mansoura District-District, Governorate and Published Estimates

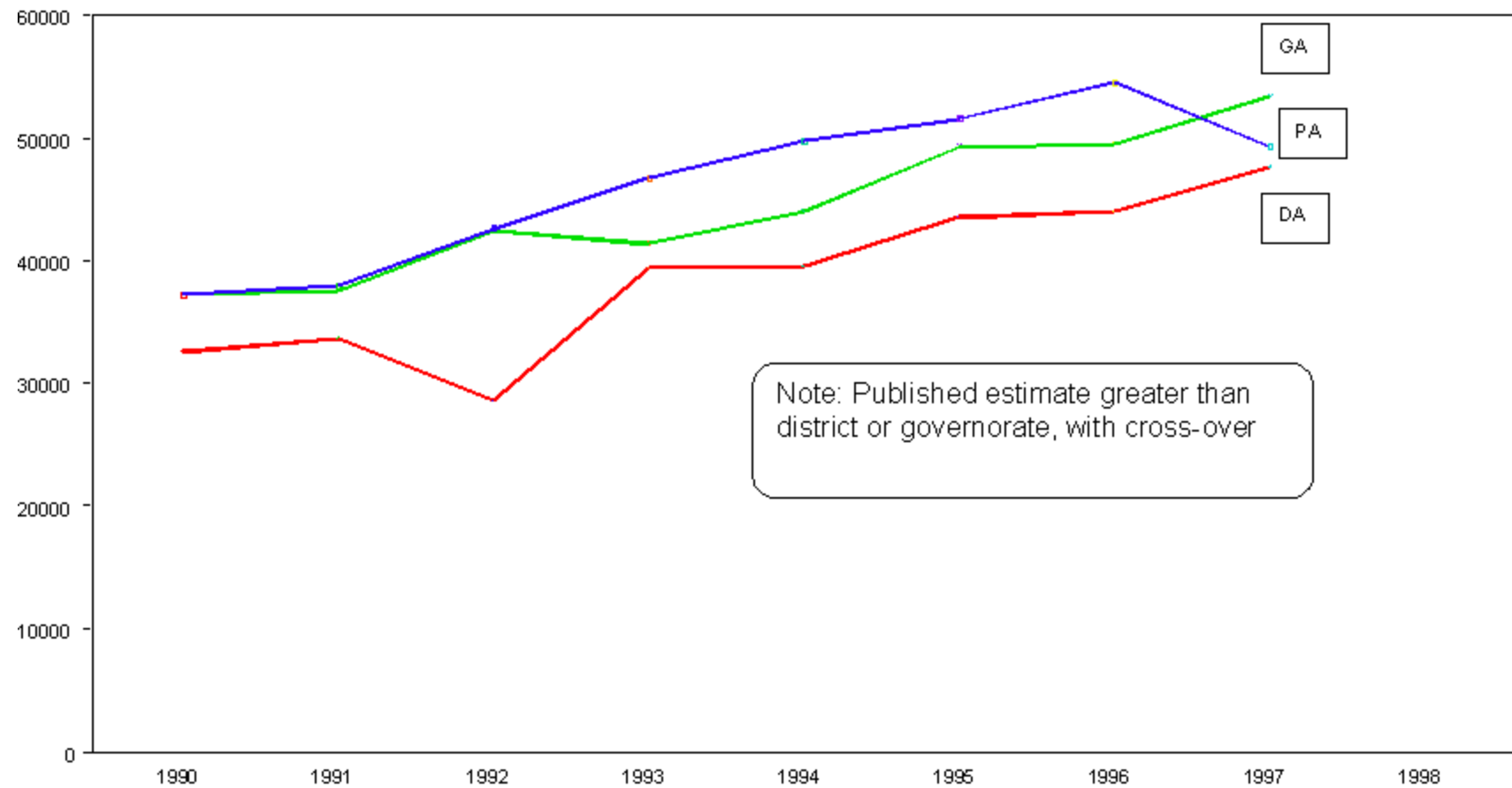
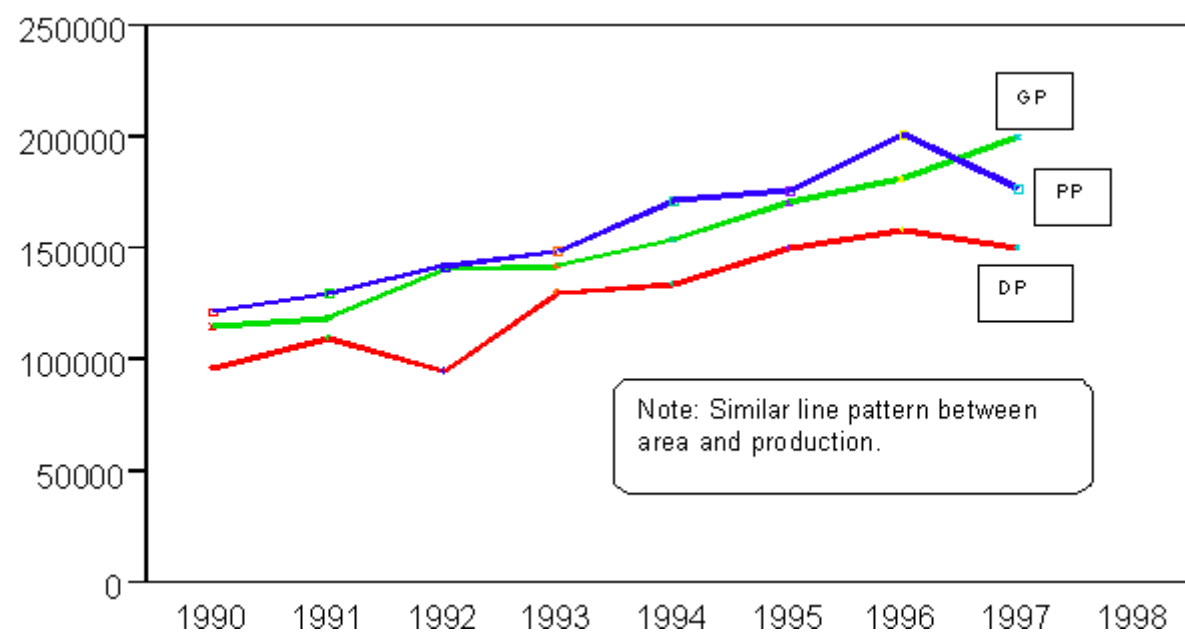


Chart 7.14: Rice Production in Dakahlia, Mansoura District -District, Governorate and Published Estimates



7.2 Estimates Moving from District to Governorate to Publication (D→G→P)

The estimates for selected districts were followed as they were transferred from the district office to the governorate office and to eventual published (official) estimates. The first five crop data analysis follows. Tables 7.1-7.5 below show the comparisons made for each crop, the percent of estimates that were changed, average change and whether this change was significant. We were told that the data were aggregated through the various levels without change. Therefore our Null Hypothesis is that there were no changes at each transfer of data:

$$H_0: M_i = M_j \text{ or } M_i - M_j = 0$$

The alternate hypothesis is:

$$H_a: M_i > M_j \text{ or } M_i - M_j > 0$$

Where i = lower level j = higher level

7.2.1 Number and Magnitude of Changes

The Tables 7.1-7.5 give the percentages of estimates that were changed as they went from district to governorate or to publication. The percent of estimates that were changed ranged from 36% to 96% with most being above 60%. This is an extremely high number of changes! Most statistical systems would expect less than 1% of estimates to be changed, and this only with detailed explanation.

The number of changes appear to be more moderate for major field crops like cotton and larger for minor crops like maize. This is probably due to less check data and more diverse use of these crops.

The column of average change gives a measure of the magnitude of change. Of course, some estimates increased and some decreased as they changed levels. On average though, areas were increased 17% and production estimates increased 14% as they went to the next higher level. All area and production comparison averages were increases except a few of the governorate to publication comparisons, which probably reflect corrections to excessive changes earlier. Fava beans were the only exception, where all average changes were decreases in estimates.

To test whether these changes were significant we used several different statistical tests:

7.2.2 Test of Paired Differences

In theory data is not changed as it goes from D→G→P. Therefore the differences DI_i , GI_i , $DI_i - PI_i$, $GI_i - PI_i$ should all be =0.

Where: $I = A, Y \text{ or } P$

i = subscript of paired observations

To test if these differences are significantly different from zero the *paired difference t-test* is an appropriate test statistic. The Tables 7.1-7.5 give the results of these tests.

In most cases the null hypotheses of no changes in the estimates as they go to higher levels are rejected. In fact most of these test results showed that the probability of a larger t-value was less than 0.0001. In other words the null hypothesis of no changes is soundly rejected!

We also did a non-parametric *Wilcoxon Signed Rank Test*, which looks at the number of positive changes and negative changes and tests these. The results of these tests were essentially the same as the paired difference tests.

7.2.3 Analysis of Variance

The *analysis of variance* (ANOVA) tests the means of treatments. In these analyses, sets of district, governorate and published data were considered as treatments and their means were tested.

The null hypothesis is that	$H_0: M_D = M_G = M_P$
The alternate hypothesis is that	$H_a: M_D \neq M_G \neq M_P$
Where D, G, P stand for level of estimate as defined above	

Of course, one wants to identify if treatment means differ more than can be expected by chance, but also identify which ones are different and estimate these treatment means, if possible. During the analysis process many different ANOVAs were considered, but three were chosen to calculate for each set of data, as they were judged to be the most appropriate. The ANOVAs chosen were governorate, governorate and districts within governorate, and districts as treatment effects. These ANOVAs were computed for each crop's area, yield and production dataset. This involved nine ANOVAs for each crop.

The results of these forty five ANOVAs are given in Tables 7.6-7.12. The test results were consistent and resoundingly rejected the null hypotheses of no difference in treatment means (means of estimates at these different levels of aggregation). Almost all of the tests had F-values so great as to have a probability of less than 0.0001 of a greater F-value.

Table 7.1: Cotton Estimate Comparisons

Variable	Comparisons Of Estimates	% of Estimates With Changes	Average Change	Ho : Mi - Mj = 0 Are changes significantly different from zero ?
Area	DA → GA	48	1516	Yes
(feddan)	DA → PA	51	2333	Yes
	GA → PA	10	2593	Yes
Yield	DY → GY	64	-0.37	Yes
(Kentar per Feddan)	DY → PY	76	-0.79	Yes
	GY → PY	61	-0.55	Yes
Production	DP → GP	67	5656	Yes
(000 Kentars)	DP → PP	82	4991	Yes
	GP → PP	74	-1225	No

The notation used in the table is as follows:

First character D = District level estimate
 G = Governorate level estimate
 P = Published (national level) estimate

Second character A = Area estimate
 Y = Yield estimate
 P = Production estimate

DA àGA Denotes comparison of District Area estimate with Governorate Area estimate

**Yes- There is a less than 5 % chance that this difference
 would occur if the null hypothesis were true.**

Yes There is a less than 1 % chance that this difference would occur if the null
 hypothesis were true.

No There is insufficient evidence to reject the null hypothesis. That is there is a
 possibility that the hypothesis is true.

Table 7.2: Wheat Estimate Comparisons

Variable	Comparisons Of Estimates	% of Estimates With Changes	Average Change	$H_0 : \mu_i - \mu_j = 0$ Are changes significantly different from zero ?
Area	DA → GA	51	2064	Yes
(feddan)	DA → PA	92	1457	Yes
	GA → PA	82	-13	No
Yield	DY → GY	49	1.48	Yes
(Ardab per Feddan)	DY → PY	81	-0.19	No
	GY → PY	82	-1.44	Yes
Production	DP → GP	55	55863	Yes
(000 Ardabs)	DP → PP	88	17155	Yes
	GP → PP	88	25012	Yes

The notation used in the table is as follows:

First character D = District level estimate
 G = Governorate level estimate
 P = Published (national level) estimate

Second character A = Area estimate
 Y = Yield estimate
 P = Production estimate

DA → GA Denotes comparison of District Area estimate with Governorate Area estimate

Yes- There is a less than 5 % chance that this difference would occur if the null hypothesis were true.

Yes There is a less than 1 % chance that this difference would occur if the null hypothesis were true.

No There is insufficient evidence to reject the null hypothesis. That is there is a possibility that the hypothesis is true.

Table 7.3: Rice Estimate Comparisons

Variable	Comparisons Of Estimates	% of Estimates With Changes	Average Change	$H_0 : \mu_i - \mu_j = 0$ Are changes significantly different from zero ?
Area	DA → GA	67	2355	Yes
(feddan)	DA → PA	91	4508	Yes
	GA → PA	69	2641	Yes
Yield	DY → GY	60	0.16	No
(ton per feddan)	DY → PY	91	0.20	Yes
	GY → PY	73	0.06	No
Production	DP → GP	69	13538	Yes
(000 tons)	DP → PP	93	20440	Yes
	GP → PP	75	8943	Yes

The notation used in the table is as follows:

First character D = District level estimate
 G = Governorate level estimate
 P = Published (national level) estimate

Second character A = Area estimate
 Y = Yield estimate
 P = Production estimate

DA → GA Denotes comparison of District Area estimate with Governorate Area estimate

Yes- There is a less than 5 % chance that this difference would occur if the null hypothesis were true.

Yes There is a less than 1 % chance that this difference would occur if the null hypothesis were true.

No There is insufficient evidence to reject the null hypothesis. That is there is a possibility that the hypothesis is true.

Table 7.4: Maize Estimate Comparisons

Variable	Comparisons Of Estimates	% of Estimates With Changes	Average Change	Ho : $M_i - M_j = 0$ Are changes significantly different from zero ?
Area	DA → GA	51	1246	Yes
(feddan)	DA → PA	60	1126	Yes
	GA → PA	36	-848	Yes
Yield	DY → GY	58	1.92	Yes
(Ardab per feddan)	DY → PY	87	1.30	Yes
	GY → PY	85	0.16	No
Production	DP → GP	60	38893	Yes
(000 Ardabs)	DP → PP	87	32523	Yes
	GP → PP	88	-4904	No

The notation used in the table is as follows:

First character D = District level estimate
 G = Governorate level estimate
 P = Published (national level) estimate

Second character A = Area estimate
 Y = Yield estimate
 P = Production estimate

DA àGA Denotes comparison of District Area estimate with Governorate Area estimate

Yes- There is a less than 5 % chance that this difference would occur if the null hypothesis were true.

Yes There is a less than 1 % chance that this difference would occur if the null hypothesis were true.

No There is insufficient evidence to reject the null hypothesis. That is there is a possibility that the hypothesis is true.

Table 7.5: Fava Bean Estimate Comparisons

Variable	Comparisons Of Estimates	% of Estimates With Changes	Average Change	Ho : $M_i - M_j = 0$ Are changes significantly different from zero ?
Area	DA → GA	36	-393	No
(feddan)	DA → PA	60	-1017	Yes
	GA → PA	67	-1857	Yes
Yield	DY → GY	51	-0.23	No
(Ardab per feddan)	DY → PY	69	-0.43	Yes-
	GY → PY	90	-0.14	No
Production	DP → GP	62	-2584	No
(000 Ardabs)	DP → PP	69	-7118	Yes
	GP → PP	96	-7047	Yes

The notation used in the table is as follows:

First character D = District level estimate
 G = Governorate level estimate
 P = Published (national level) estimate

Second character A = Area estimate
 Y = Yield estimate
 P = Production estimate

DA à GA Denotes comparison of District Area estimate with Governorate Area estimate

Yes- There is a less than 5 % chance that this difference would occur if the null hypothesis were true.

Yes There is a less than 1 % chance that this difference would occur if the null hypothesis were true.

No There is insufficient evidence to reject the null hypothesis. That is there is a possibility that the hypothesis is true.

Table 7.6: Cotton – Analysis of Variance (ANOVA) Results

Area Estimates

Source	d.f.	F ratio	Prob. > F	R ²
Governorate	3	19.707	<0.0001	0.193
Governorate	3	58.201	<0.0001	0.778
District w/in Gov.	8	78.852	<0.0001	
District	11	76.275	<0.0001	0.778

Yield Estimates

Source	d.f.	F ratio	Prob. > F	R ²
Governorate	3	12.490	<0.0001	0.133
Governorate	3	2.000	0.0044	0.209
District w/in Gov.	8	12.950	<0.0001	
District	11	5.715	<0.0001	0.209

Production Estimates

Source	d.f.	F ratio	Prob. > F	R ²
Governorate	3	24.250	<0.0001	0.230
Governorate	3	87.000	<0.0001	0.833
District w/in Gov.	8	107.000	<0.0001	
District	11	107.744	<0.0001	0.833

Table 7.7: Wheat – Analysis of Variance (ANOVA) Results

Area Estimates

Source	d.f.	F ratio	Prob. > F	R ²
Governorate	3	10.235	<0.0001	0.103
Governorate	3	52.062	<0.0001	0.834
District w/in Gov.	8	143.643	<0.0001	
District	11	119.146	<0.0001	0.834

Yield Estimates

Source	d.f.	F ratio	Prob. > F	R ²
Governorate	3	21.143	<0.001	0.196
Governorate	3	21.303	<0.0001	0.234
District w/in Gov.	8	1.543	0.1429	
District	11	6.985	<0.0001	0.234

Production Estimates

Source	d.f.	F ratio	Prob. > F	R ²
Governorate	3	15.463	<0.0001	0.148
Governorate	3	43.002	<0.0001	0.707
District w/in Gov.	8	61.976	<0.0001	
District	11	56.966	<0.0001	0.707

Table 7.8: Rice – Analysis of Variance (ANOVA) Results

Area Estimates

Source	d.f.	F ratio	Prob. > F	R ²
Governorate	2	198.900	<0.0001	0.741
Governorate	2	614.508	<0.0001	0.925
District w/in Gov.	6	54.843	<0.0001	
District	8	206.425	<0.0001	0.925

Yield Estimates

Source	d.f.	F ratio	Prob. > F	R ²
Governorate	2	2162.000	<0.0001	0.969
Governorate	2	1756.961	<0.0044	0.973
District w/in Gov.	6	3.823	<0.0001	
District	8	609.222	<0.0001	0.973

Production Estimates

Source	d.f.	F ratio	Prob. > F	R ²
Governorate	2	155.269	<0.0001	0.691
Governorate	2	342.114	<0.0001	0.874
District w/in Gov.	6	32.273	<0.0001	
District	8	115.422	<0.0001	0.874

Table 7.9: Maize – Analysis of Variance (ANOVA) Results

Area Estimates

Source	d.f.	F ratio	Prob. > F	R ²
Governorate	3	36.422	<0.0001	0.305
Governorate	3	262.737	<0.0001	0.899
District w/in Gov.	8	177.935	<0.0001	
District	11	195.080	<0.0001	0.899

Yield Estimates

Source	d.f.	F ratio	Prob. > F	R ²
Governorate	3	96.849	<0.0001	0.539
Governorate	3	115.858	<0.0001	0.619
District w/in Gov.	8	6.337	<0.0001	
District	11	35.551	<0.0001	0.619

Production Estimates

Source	d.f.	F ratio	Prob. > F	R ²
Governorate	3	40.743	<0.0001	0.329
Governorate	3	232.873	<0.0001	0.872
District w/in Gov.	8	128.136	<0.0001	
District	11	149.690	<0.0001	0.872

Table 7.10: Fava Beans – Analysis of Variance (ANOVA) Results

Area Estimates

Source	d.f.	F ratio	Prob. > F	R ²
Governorate	2	17.661	<0.0001	0.167
Governorate	2	21.613	<0.0001	0.341
District w/in Gov.	6	7.449	<0.0001	
District	8	10.973	<0.0001	0.341

Yield Estimates

Source	d.f.	F ratio	Prob. > F	R ²
Governorate	2	14.451	<0.0001	0.141
Governorate	2	14.808	<0.0001	0.173
District w/in Gov.	6	1.109	0.359	
District	8	4.458	<0.0001	0.173

Production Estimates

Source	d.f.	F ratio	Prob. > F	R ²
Governorate	2	32.259	<0.0001	0.268
Governorate	2	35.051	<0.0001	0.359
District w/in Gov.	6	3.990	0.0009	
District	8	11.880	<0.0001	0.359

7.3 Estimates Moving from Cooperative (Village) to District to Governorate (C→D→G)

Review of village level data showed an unusually large number of obvious errors like decimals misplaced or omissions of data values. Many transcription errors were found and some other suspect data relationships were noted. (Some of the same problems were noted in the D→G→P analysis, but the number of errors was much smaller.) This is mentioned here to emphasize the need for training of staff and the implementation of data handling procedures that include checks on data quality. Prior to analysis, obvious errors that would distort the analysis were corrected by the MVE team.

The comparison of data patterns among governorates and districts indicate that slightly different procedures were used to handle village data. The most complete set of village data was found at different administrative levels depending on governorate. The data for a few villages in one governorate had such uniform area, yield and production that they looked suspicious. It appears that constants or technical coefficients have been used instead of current estimates. If so, the village estimation may have actually taken place at the district or higher level without consultation with village agents.

There were two villages within each of the three districts in each of four governorates. Analysis was done for the same five crops as in the above section. Generally the C→D→G data were more sparse than the D→G→P data, but was still adequate for the analysis procedures.

7.3.1 Number and Magnitude of Changes

We follow the logic of the above section. The Tables 7.11-7.15 give the percentages of estimates that were changed as they went from cooperative to district to governorate levels. The percent of estimates that were changed ranged from 26% to 100% with most being above 50%. As stated earlier this is an extremely high number of changes!

The column of average change gives a measure of the magnitude of change. Since we are reviewing village area and production the magnitude of changes is much less than the previous district level analyses. These changes are still important, nevertheless. On average, area changes were about 7% and production changes were about 11%. Interestingly, there are about an equal number of decreases as increases, which is in sharp contrast to what we found at the district level, where most of the changes were increases.

7.3.2 Test of paired differences

In theory data is not changed as it goes from C→D→G. Therefore the differences $CI_i - DI_i$, $CI_i - GI_i$, $DI_i - GI_i$ should all be =0.

Where: $I = A, Y \text{ or } P$
 $i = \text{subscript of paired observations}$

To test if these differences are significantly different from zero the *paired difference t-test* is an appropriate test statistic. The tables 7.11-7.15 gives the results of these tests in the last column.

Note that over half of the tests did show some level of significant data changes, but many were non-significant. There was no consistent pattern of changes up or down through levels of estimates or different crops.

We also did a non-parametric *Wilcoxon Signed Rank Test*, which looks at the number of positive changes and negative changes and tests these numbers. The results of these tests were essentially the same as the paired difference tests.

7.3.3 Analysis of Variance

The *analysis of variance* (ANOVA) tests the means of treatments. In these analyses, sets of village, district and governorate data were considered as treatments and their means were tested.

The null hypothesis is that	$H_0: M_C = M_D = M_G$
The alternate hypothesis is that	$H_a: M_C \neq M_D \neq M_G$
Where C, D, G stand for level of estimate as defined above	

Of course, one wants to identify if treatment means differ, but also identify which ones are different and estimate these treatment means, if possible. During the analysis process many different ANOVAs were considered, but five were chosen to calculate for each set of data. The two ANOVAs judged to be the most appropriate were village effects and the nested (governorate - district within governorate village within district and governorate). The three ANOVAs governorate, governorate and districts within governorate, and districts as treatment effects were also calculated for comparability with the previous section analyses. All ANOVAs were computed for each crop's area, yield and production dataset. This involved fifteen ANOVAs for each crop.

The results of these seventy five ANOVAs are given in Tables 7.16-7.20. The test results were consistent and resoundingly rejected the null hypotheses of no difference in treatment means (means of estimates at these different levels of aggregation). Most of the tests had F-values so great as to have a probability of less than 0.0001 of a greater F-value. Those that had higher probabilities were still low enough to be significant. Only two minor test results were not significant.

These analyses shows that:

- Changes are being made in the data as it moves through all levels from village to district to governorate to national (publication). Statistical tests indicate that changes are significant at most levels and crops.
- Changes are most pronounced as data moves from D→G→P and tend to increase area and production estimates.

- A valuable source, village extension agent data, is not being adequately acknowledged or utilized!
- Training in proper data handling is needed at all levels of the process and data quality checks need to be instituted.

Table 7.11: Cotton Estimate Comparisons – Village Level

Variable	Comparisons Of Estimates	% of Estimates With Changes	Average Change	$H_0 : \mu_i - \mu_j = 0$ Are changes significantly different from zero ?
Area	CA → DA	57	-54	Yes
(feddan)	CA → GA	64	-24	Yes-
	DA → GA	28	5	No
Yield	CY → DY	57	0.33	No
(Kentar per Feddan)	CY → GY	62	-0.19	No
	DY → GY	50	-0.62	Yes
Production	CP → DP	64	-388	Yes-
(000 Kentars)	CP → GP	66	-175	No
	DP → GP	43	2	Yes-

The notation used in the table is as follows:

First character C = Cooperative or Village level estimates
 D = District level estimate
 G = Governorate level estimate
 Second character A = Area estimate
 Y = Yield estimate
 P = Production estimate

DA à GA Denotes comparison of District Area estimate with Governorate Area estimate

Yes- There is a less than 5 % chance that this difference would occur if the null hypothesis were true.

Yes There is a less than 1 % chance that this difference would occur if the null hypothesis were true.

No There is insufficient evidence to reject the null hypothesis. That is there is a possibility that the hypothesis is true.

Table 7.12: Wheat Estimate Comparisons– Village Level

Variable	Comparisons Of Estimates	% of Estimates With Changes	Average Change	$H_0 : M_i - M_j = 0$ Are changes significantly different from zero ?
Area	CA → DA	61	-10.28	No
(feddan)	CA → GA	74	3.5	No
	DA → GA	26	-19.37	No
Yield	CY → DY	58	3.50	Yes
(Ardabs per feddan)	CY → GY	64	4.12	Yes
	DY → GY	41	0.67	Yes-
Production	CP → DP	61	2179	Yes
(000 Ardabs)	CP → GP	69	2185	Yes
	DP → GP	46	87	No

The notation used in the table is as follows:

First character C = Cooperative or Village level estimates

D = District level estimate

G = Governorate level estimate

Second character A = Area estimate

Y = Yield estimate

P = Production estimate

DA à GA Denotes comparison of District Area estimate with Governorate Area estimate

Yes- There is a less than 5 % chance that this difference would occur if the null hypothesis were true.

Yes There is a less than 1 % chance that this difference would occur if the null hypothesis were true.

No There is insufficient evidence to reject the null hypothesis. That is there is a possibility that the hypothesis is true.

Table 7.13: Rice Estimate Comparisons– Village Level

Variable	Comparisons Of Estimates	% of Estimates With Changes	Average Change	$H_0 : \mu_i - \mu_j = 0$ Are changes significantly different from zero ?
Area	CA → DA	77	26.30	No
(feddan)	CA → GA	92	58.30	Yes-
	DA → GA	50	35.57	No
Yield	CY → DY	73	0.42	Yes
(ton per feddan)	CY → GY	90	0.47	Yes
	DY → GY	60	0.05	No
Production	CP → DP	73	327	Yes
(000 tons)	CP → GP	100	468	Yes
	DP → GP	67	160	Yes-

The notation used in the table is as follows:

First character C = Cooperative or Village level estimates
 D = District level estimate
 G = Governorate level estimate
 Second character A = Area estimate
 Y = Yield estimate
 P = Production estimate

DA à GA Denotes comparison of District Area estimate with Governorate Area estimate.

Yes- There is a less than 5 % chance that this difference would occur if the null hypothesis were true.

Yes There is a less than 1 % chance that this difference would occur if the null hypothesis were true.

No There is insufficient evidence to reject the null hypothesis. That is there is a possibility that the hypothesis is true.

Table 7.14: Maize Estimate Comparisons– Village Level

Variable	Comparisons Of Estimates	% of Estimates With Changes	Average Change	$H_0 : M_i - M_j = 0$ Are changes significantly different from zero ?
Area	CA → DA	50	87.15	No
(feddan)	CA → GA	58	-57.78	No
	DA → GA	21	-83.77	Yes
Yield	CY → DY	59	1.94	Yes
(Ardabs per feddan)	CY → GY	66	3.44	Yes
	DY → GY	27	1.69	Yes
Production	CP → DP	59	1730	No
(000 Ardabs)	CP → GP	68	-772	No
	DP → GP	32	-803	Yes-

The notation used in the table is as follows:

First character C = Cooperative or Village level estimates

D = District level estimate

G = Governorate level estimate

Second character A = Area estimate

Y = Yield estimate

P = Production estimate

DA à GA Denotes comparison of District Area estimate with Governorate Area estimate

Yes- There is a less than 5 % chance that this difference would occur if the null hypothesis were true.

Yes There is a less than 1 % chance that this difference would occur if the null hypothesis were true.

No There is insufficient evidence to reject the null hypothesis. That is there is a possibility that the hypothesis is true.

Table 7.15: Fava Bean Estimate Comparisons– Village Level

Variable	Comparisons Of Estimates	% of Estimates With Changes	Average Change	$H_0 : \mu_i - \mu_j = 0$ Are changes significantly different from zero ?
Area	CA → DA	54	-1.02	No
(feddan)	CA → GA	61	-25.89	Yes
	DA → GA	24	-41.69	Yes-
Yield	CY → DY	51	-0.33	No
(ardabs per feddan)	CY → GY	59	-0.04	No
	DY → GY	30	0.77	Yes-
Production	CP → DP	57	-73.10	No
(000 ardabs)	CP → GP	62	-212.60	Yes-
	DP → GP	33	-122.64	No

The notation used in the table is as follows:

First character C = Cooperative or Village level estimates

D = District level estimate

G = Governorate level estimate

Second character A = Area estimate

Y = Yield estimate

P = Production estimate

DA à GA Denotes comparison of District Area estimate with Governorate Area estimate

Yes- There is a less than 5 % chance that this difference would occur if the null hypothesis were true.

Yes There is a less than 1 % chance that this difference would occur if the null hypothesis were true.

No There is insufficient evidence to reject the null hypothesis. That is there is a possibility that the hypothesis is true.

Table 7.16: Cotton – Analysis of Variance (ANOVA) Results

Area Estimates – Village Level

Source	d.f.	F ratio	Prob. > F	R ²
Village	23	191.860	<0.0001	0.918
Governorate	3	211.064	<0.0001	0.918
Dist. w/in Gov.	8	222.065	<0.0001	
Vil w/in Dist, Gov	12	151.977	<0.0001	
District	11	43.176	<0.0001	0.539
Governorate	3	40.489	<0.0001	0.539
District w/in Gov.	8	39.403	<0.0001	
Governorate	3	30.590	<0.0001	0.181

Yield Estimates– Village Level

Source	d.f.	F ratio	Prob. > F	R ²
Village	23	4.250	<0.0001	0.217
Governorate	3	4.573	0.0037	0.217
Dist. w/in Gov.	8	6.108	<0.0001	
Vil w/in Dist, Gov	12	2.716	0.0016	
District	11	5.608	<0.0001	0.144
Governorate	3	4.238	0.0058	0.144
District w/in Gov.	8	5.849	<0.0001	
Governorate	3	4.499	0.0041	0.035

Production Estimates– Village Level

Source	d.f.	F ratio	Prob. > F	R ²
Village	23	120.369	<0.0001	0.887
Governorate	3	106.070	<0.0001	0.887
Dist. w/in Gov.	8	123.357	<0.0001	
Vil w/in Dist, Gov	12	89.240	<0.0001	
District	11	39.641	<0.0001	0.544
Governorate	3	31.208	<0.0001	0.544
District w/in Gov.	8	36.165	<0.0001	
Governorate	3	27.915	<0.0001	0.183

Table 7.17: Wheat – Analysis of Variance (ANOVA) Results

Area Estimates – Village Level

Source	d.f.	F ratio	Prob. > F	R ²
Village	23	254.126	<0.0001	0.936
Governorate	3	608.328	<0.0001	0.936
Dist. w/in Gov.	8	75.066	<0.0001	
Vil w/in Dist, Gov	12	253.102	<0.0001	
District	11	30.660	<0.0001	0.450
Governorate	3	89.580	<0.0001	0.450
District w/in Gov.	8	10.813	<0.0001	
Governorate	3	70.449	<0.0001	0.334

Yield Estimates– Village Level

Source	d.f.	F ratio	Prob. > F	R ²
Village	23	3.806	<0.0001	0.187
Governorate	3	13.969	<0.0001	0.187
Dist. w/in Gov.	8	3.282	0.0012	
Vil w/in Dist, Gov	12	1.133	0.3315	
District	11	6.694	<0.0001	0.158
Governorate	3	13.782	<0.0001	0.158
District w/in Gov.	8	3.216	0.0015	
Governorate	3	15.293	<0.0001	0.103

Production Estimates– Village Level

Source	d.f.	F ratio	Prob. > F	R ²
Village	23	156.458	<0.0001	0.904
Governorate	3	399.177	<0.0001	0.904
Dist. w/in Gov.	8	38.739	<0.0001	
Vil w/in Dist, Gov	12	152.295	<0.0001	
District	11	28.649	<0.0001	0.445
Governorate	3	85.431	<0.0001	0.445
District w/in Gov.	8	9.679	<0.0001	
Governorate	3	67.543	<0.0001	0.336

Table 7.18: Rice – Analysis of Variance (ANOVA) Results

Area Estimates – Village Level

Source	d.f.	F ratio	Prob. > F	R ²
Village	11	101.426	<0.0001	0.862
Governorate	1	31.410	<0.0001	0.862
Dist. w/in Gov.	4	178.566	<0.0001	
Vil w/in Dist, Gov	6	81.692	<0.0001	
District	5	34.452	<0.0001	0.484
Governorate	1	12.139	0.0006	0.484
District w/in Gov.	4	42.990	<0.0001	
Governorate	1	0.153	0.6960	0.0008

Yield Estimates– Village Level

Source	d.f.	F ratio	Prob. > F	R ²
Village	11	5.995	<0.0001	0.291
Governorate	1	10.004	0.0019	0.291
Dist. w/in Gov.	4	8.996	<0.0001	
Vil w/in Dist, Gov	6	4.264	0.0005	
District	5	7.224	<0.0001	0.178
Governorate	1	8.579	0.0040	0.178
District w/in Gov.	4	8.143	<0.0001	
Governorate	1	3.041	0.0830	0.017

Production Estimates– Village Level

Source	d.f.	F ratio	Prob. > F	R ²
Village	11	56.387	<0.0001	0.794
Governorate	1	15.389	0.0001	0.794
Dist. w/in Gov.	4	93.666	<0.0001	
Vil w/in Dist, Gov	6	38.801	<0.0001	
District	5	32.861	<0.0001	0.496
Governorate	1	6.284	0.0131	0.496
District w/in Gov.	4	40.754	<0.0001	
Governorate	1	0.669	0.4150	0.004

Table 7.19: Maize – Analysis of Variance (ANOVA) Results

Area Estimates – Village Level

Source	d.f.	F ratio	Prob. > F	R ²
Village	21	62.299	<0.0001	0.813
Governorate	3	99.587	<0.0001	0.813
Dist. w/in Gov.	7	60.037	<0.0001	
Vil w/in Dist, Gov	11	35.443	<0.0001	
District	10	41.475	<0.0001	0.571
Governorate	3	58.547	<0.0001	0.571
District w/in Gov.	7	38.552	<0.0001	
Governorate	3	26.476	<0.0001	0.199

Yield Estimates– Village Level

Source	d.f.	F ratio	Prob. > F	R ²
Village	21	12.916	<0.0001	0.482
Governorate	3	75.690	<0.0001	0.482
Dist. w/in Gov.	7	3.888	0.0005	
Vil w/in Dist, Gov	11	2.439	0.0064	
District	10	23.225	<0.0001	0.435
Governorate	3	71.881	<0.0001	0.435
District w/in Gov.	7	3.755	0.0006	
Governorate	3	64.621	<0.0001	0.386

Production Estimates– Village Level

Source	d.f.	F ratio	Prob. > F	R ²
Village	21	53.768	<0.0001	0.795
Governorate	3	94.962	<0.0001	0.795
Dist. w/in Gov.	7	52.492	<0.0001	
Vil w/in Dist, Gov	11	30.626	<0.0001	
District	10	38.106	<0.0001	0.558
Governorate	3	55.966	<0.0001	0.558
District w/in Gov.	7	34.636	<0.0001	
Governorate	3	26.221	<0.0001	0.203

Table 7.20: Fava Bean – Analysis of Variance (ANOVA) Results

Area Estimates – Village Level

Source	d.f.	F ratio	Prob. > F	R ²
Village	17	43.363	<0.0001	0.694
Governorate	2	83.703	<0.0001	0.694
Dist. w/in Gov.	6	42.362	<0.0001	
Vil w/in Dist, Gov	9	25.537	<0.0001	
District	8	38.177	<0.0001	0.478
Governorate	2	71.141	<0.0001	0.478
District w/in Gov.	6	29.394	<0.0001	
Governorate	2	42.988	<0.0001	0.202

Yield Estimates– Village Level

Source	d.f.	F ratio	Prob. > F	R ²
Village	17	4.952	<0.0001	0.210
Governorate	2	13.488	<0.0001	0.210
Dist. w/in Gov.	6	4.968	<0.0001	
Vil w/in Dist, Gov	9	2.877	0.0028	
District	8	6.925	<0.0001	0.146
Governorate	2	12.744	<0.0001	0.146
District w/in Gov.	6	4.765	0.0001	
Governorate	2	12.551	<0.0001	0.070

Production Estimates– Village Level

Source	d.f.	F ratio	Prob. > F	R ²
Village	17	60.212	<0.0001	0.764
Governorate	2	108.267	<0.0001	0.764
Dist. w/in Gov.	6	54.256	<0.0001	
Vil w/in Dist, Gov	9	38.255	<0.0001	
District	8	41.794	<0.0001	0.507
Governorate	2	74.512	<0.0001	0.507
District w/in Gov.	6	33.899	<0.0001	
Governorate	2	41.019	<0.0001	0.199

7.4 Cost of Production Data

As it has been mentioned earlier (in Chapter 5) that in Egypt there are a large number of small farms, which make up an important part of the agricultural economy. Their costs, returns and economic condition are very important to agricultural policy and impact considerations. Data on their operations are critical and should not be ignored or under-represented in statistics.

7.4.1 Published Data of Total Cost of Production

There are estimates of *cost of production*, *farm gate prices*, *labor and wage rates* available for governorates. However, our investigation found that most of these were very subjective and covered a very narrow segment of the sector (see Annex C). Many of the village extension agents had such data, but stated that they were not asked for it. The governorate level data was often gathered at the ministry office from staff who were also farmers. Their costs are hardly those of the farmer on less than one feddan.

As part of our work, we tested gathering *cost of production* and *farm gate prices* from three farmers in each village. We found that their prices were very different from the published governorate prices. They also were different from village to village and from district to district (see table 7-21).

When conducting the time series analysis (Chapter 6) it was very unexpected to find that the time series data of total cost lay on a straight line (see charts 7-15 to 7-17). This unusual result held true for total cost data of wheat, rice, maize, and cotton. As it is already well known that total costs of production are obviously affected by: (1) The prices and quantities of the inputs used in the production process, (2) The level of output produced, and (3) The state of technical knowledge used in production. The directions and variations in the above three factors are not the same. Therefore, we can - without hesitation - say that the total cost time series data are too uniform and thus likely not accurate. Again, some interference is indicated during the collection and reporting processes of these data.

7.4.2 Potential improvement of Costs of Production Data Quality

The main findings of both of the field survey and time series analysis showed clearly the necessity of establishing a better system of cost and price gathering. A village level survey could be instituted to give much more representative and reliable information. With some training and support extension agents could provide good cost of production, farm gate prices, estimate volumes produced, and many other types of data. This can be done through proper sampling. Thus a system should be devised to make sure that these are given proper representation in the coverage. The sampling procedure which has been applied in this study can be adopted or adapted and followed with a larger sample size.

Table 7.21: Variable Costs, Rent, and Total Cost of Production for Major Field Crops in the Selected Governorates as in 1997

Chart 7.15: Cost of Production -Wheat, Beni Suef Governorate

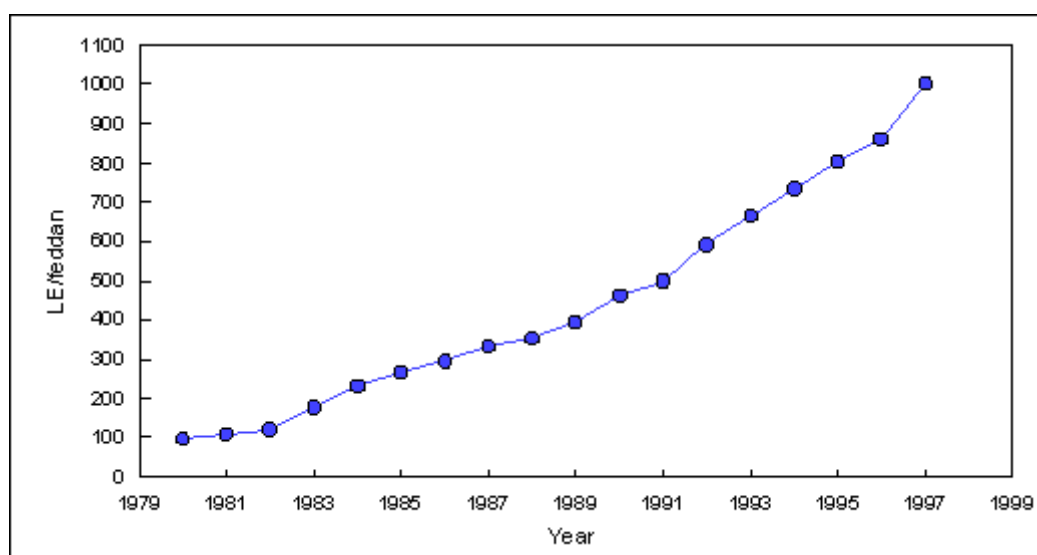


Chart 7.16: Cost of Production -Cotton, Beni Suef Governorate

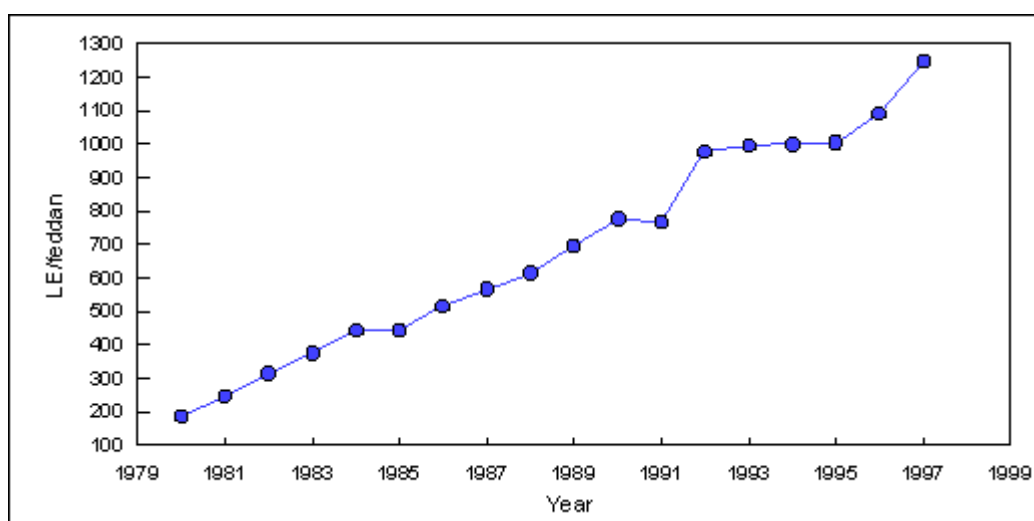
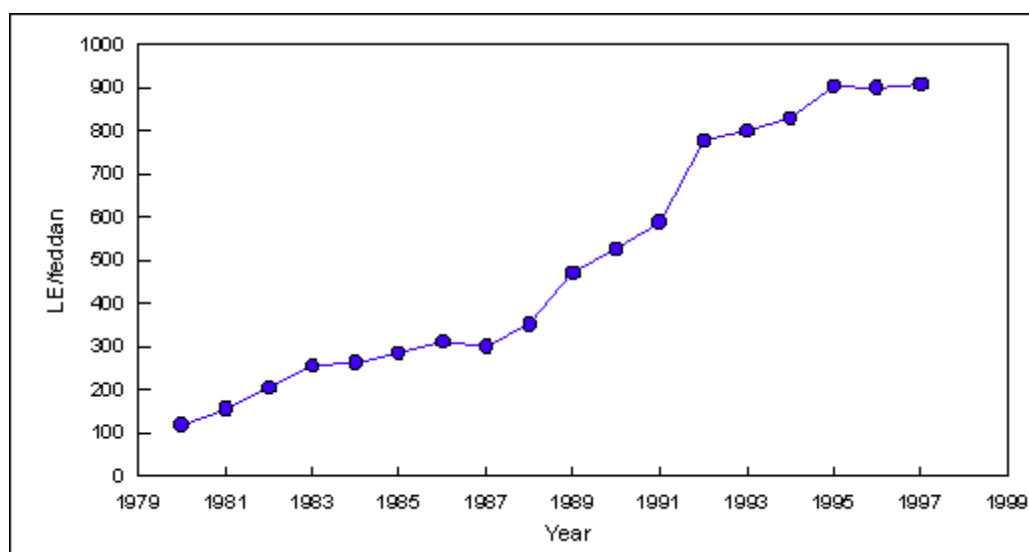


Chart 7.17: Cost of Production -Maize, Beni Suf Governorate



8. CONCLUSIONS AND RECOMMENDATIONS

Others have looked at the data collection process and identified the theoretical versus actual procedures. The current study chose to examine these issues in more depth and detail than past studies. Some of the work accomplished had never been done before, including to:

- Follow the data through all levels from the source villages, through district, governorate offices to the national office. Determine if changes are made, where, and why.
- Study in detail the data gathering process, identify problems and constraints to good data handling at each level, including the education and training of the staff and technical direction given by the national level.
- Create an electronic database at the district level for the four governorates in which we were carrying out our investigation.
- Conduct time series analysis on the MALR data series, to see what the data can tell about the process.
- Test possible methods that could be used to collect data that are currently not available, but needed for modern analysis and estimation procedures.
- Look in detail at the data available and whether it is accessible and reliable for modern modeling and analysis programs.

8.1 General Conclusions

This study focused on the major field crops. Cotton data did appear to be of better quality, due to the emphasis placed on it. Other crops did not fare as well. One can say that the area data for major crops should be fairly reliable if obtained at the village level, before changes are made. The yield estimates of the sampling offices are felt to be good. If these crop yield estimates are applied to valid area estimates, then reasonably good production estimates may be derived. All other data, including cost and return data, are suspect.

8.1.1 Operation of Data Collection System

- The statistical, sampling and area measurement office staffs are to be commended for trying to carry out their mandates under often very difficult conditions with little or no administrative support. Chapter 5, and especially section 5.1.4, enumerates the problems and constraints faced by the statistical organizations.
- At all levels, the staff are doing work for which they neither have training nor are given any support or recognition. The statistical work is considered of minor importance by the village extension staff, especially since they receive no incentive payment for it.

- The current statistical system is inadequate to meet current and future information and data analysis requirements.
- Most offices have plans for quality control procedures, but admit that these often have to be bypassed due to equipment or budgetary constraints. During this review, several small problems were identified that affect data quality, but they could be eliminated with a little training and support.
- The manipulation of data as it travels up through higher levels is very prevalent and thus accuracy is doubtful. There seems to be a governmental disincentive to accurate information. Many staff tell of their data being changed so that higher level government officials could receive awards or recognition for high yield or production.
- The Egyptian Survey Authority does much work and is somewhat independent of the Agricultural Affairs work, but there is also duplication of effort between these two agencies. Significant savings could probably be realized if there were a restructuring and improvement of the area estimation process.

Most of these problems have negative effects on the quality of agricultural data.

8.1.2 Data Quality, Availability and Utility

- The time lag between data gathering and publication is too great to meet modern requirements. Some data is not available early or frequently enough to meet data users needs. Proper data handling techniques and use of computers would make information more timely.
- In spite of the conditions in the village cooperatives, the extension agent records and maps do contain excellent *area* estimates. Agents have good knowledge of yields and production and other farmer related information. With some training and support they could provide good cost of production, farm gate prices, estimate volumes produced, and many other types of data.
- The best *yield* data comes from the sampling offices, which conduct crop cutting surveys at harvest time. Support should be given to verify their procedures and sample representativeness, while upgrading their equipment.
- Many data gaps exist. Items of data needed for decisions under a market driven economy are missing. A comparison of existing agricultural data in Table 5.1 with the items needed in Table 5.2 makes the data gaps very obvious. A few items are available at the governorate or national level, but are needed at much lower levels, i.e. per feddan, per farm, by farm type or geographical region.
- There are estimates of *cost of production, farm gate prices, labor and wage rates* available for governorates. However, our investigation found that most of these were very subjective and covered a very narrow segment of the sector. They lack

completeness of coverage and are too thin to be reliable. Many of the village extension agents do have such data, but are not asked to supply it. If obtained from the agents, the coverage and reliability would be greatly improved.

- As part of our work, we tested gathering *cost of production* and *farm gate prices* from three farmers in each village. We found that farmer reported prices were very different from the published governorate prices. They also were different from village to village. Clearly a better system of cost and price gathering is necessary.
- The logical implications for organizations doing impact, economic and policy analyses is that they will have to design and execute their own surveys, at least until a national statistical program can be established.
- The time series analyses which are designed to let the data tell about itself, were especially interesting. They clearly showed that in the past years the data on *area*, *yield*, *production* and *costs* were unusually stable. This indicates that normal variation expected did not occur, i.e. there was some controlling process affecting the data.
- Accessing specific information is often very difficult. Some basic and intermediate data is only kept for a year or two due to storage space or records maintenance problems. Most data is on handwritten data forms and data management is cumbersome.
- The database development work showed some promise as a potential source of data for future analysis. Work required entering data from publications, summary sheets and many dispersed sources. Some data were no longer available, as noted above. There were problems converting data to standardized units of measurement. Computer consistency checks enabled catching and correcting some errors. Further work is required to get the data files in a more user friendly format.

8.2 Recommendations

- While the data quality may not be very good at present, the prospects for the future could be bright. There is an existing infrastructure that could be built upon to develop a much improved statistical organization. Section 8.3.1 discusses this need and gives some recommendations for improving the infrastructure. We recommend using every opportunity to encourage and support infrastructure improvement of the statistical community.
- The extension agents, with a little support and training could be a valuable source of current and reliable data. The GOE should cultivate this source of data through recognition, training and support of the extension agent's statistical work. Section 8.3.2 offers training ideas.
- The district, governorate and national staff, with proper training and support, could be safe conduits to aggregated and disseminated data. Once clear lines of authority and

responsibility are designated and proper technical supervision is established, the organization should function well. Section 8.3.2 offers a plan for training which could be given at all levels, from extension agent up through national offices. This training should be pursued and support given to improvement of the statistical offices.

- It would be advantageous to start the objective yield work again to give forecasts of yields early in the season. These could be helpful for early warning of impending shortages or surpluses in production. The sampling offices should be given support to improve their current work and to begin pilot objective yield work.
- With the changes being made as data moves to higher levels of aggregation, the best place to gather the data is at the village level. Since much analysis requires detailed farmer operation and family data, it is logical that data gathering be concentrated at the village level. For example, cost of production, farmgate prices and other economic data should be gathered at the farmer level, weighted and aggregated to whatever level necessary. This would give much more representative and reliable information. Our work has demonstrated that gathering farm level data is possible. We recommend pilot work be explored to gather general cost and returns data from farmers at the village level.
- Recognizing the lack of data available at low enough levels of aggregation and for required time periods, it is recommended that representative village sample surveys be designed and carried out to provide the information for impact, economic and policy analysis.

8.3 Suggested Improvements to Statistical Organizations

8.3.1 Restructuring Statistical Organizations

As mentioned in chapter 1, the structure of farming has totally changed over the past 50 years. It has gone from a small number of very large holdings, managed by land owners, with little government control; through break-up of holdings and total government control; to very large number of small holdings, independently managed, with the Government taking on a role of advisor and assistant. Along with these structural changes the need for information and statistics has changed radically: from little need for information; to accounting type data to assist a few decision makers; to very detailed and complex data on all aspects of the agricultural economy, to assist millions of decision makers.

Each decision made by a farmer, trader, importer/exporter or government policy maker needs good information. Presently little data is available, so each has to gather his own or make his decision in an information vacuum. Having a statistical organization disseminating accurate information freely makes this myriad of information gathering activities unnecessary, thus freeing up efforts for more productive work.

The current and future needs for information are increasing at an exponential rate while the current capability to provide the information is nearly static. The Government must act quickly to create an environment for accurate, timely statistics to develop.

The GOE should implement specific structural changes to improve the agricultural statistics program. These should include:

- Delegation of specific responsibility and authority to each office and level of government.
- Decrees, if necessary, to assure protection of the statistical organization from external influence and the confidentiality of farmer disclosed information from exposure to taxing or regulatory agencies.

Concurrent with the above changes, technical training to staff will be required to make the transition effective. Some support with equipment and supplies will probably also be advantageous.

8.3.2 Statistical Training Program

This program is designed for training statistical staff working at various levels in statistical offices. The program provides for comprehensive training, within certain time limits, covering a succession of courses. For general training, all five courses should be completed in sequence. This affords coverage of all important aspects of statistical work.

However, each course can be offered independently to suit specific training needs. Also, course organization allows for modifications that may be found necessary to suit the specific personnel to be trained, such as condensing or expanding specific topics. For a particular investigation, such as an agricultural census or sample survey, these courses would be suitably expanded and supplemented to provide for their special skill.

This program starts with manual processing procedures, since this is the current organizational capability. However, training will move through manual processing to prepare for electronic data processing facilities. Each level of training will stress techniques to reduce errors and improve accuracy.

The training program is divided into five courses. The instruction is given through lectures, discussions, demonstrations, exercises and visits for each course. The courses provide knowledge on the basic concept of statistical methods and the major aspects of statistical work in government (ministry, governorates, districts, villages).

- | | |
|--------------------|---|
| Course (1) : | Provides an introduction to the nature and scope of statistics, with emphasis on official statistics and the role of international organizations in statistical activities. |
| Course (2) : | Covers principles and methods of statistics and their application to the data needs of government. |
| Course (3) & (4) : | Covers the details of statistical operations, specifically data collection methods and data processing, tabulation and presentation. |

Course (5) : Provides an overview of statistical systems, coordination and standards.

The emphasis of the training is on the practical matters. Instruction should be mainly through demonstrations, discussions and assignments involving the use of actual data, worksheets or forms and case studies. In addition, planned visits to statistical offices or other places of statistical activity may be conducted during some courses. A small scale statistical project may be undertaken to provide the trainees with additional work experience on the various aspects of statistical work.

Each course can be completed in 25 hours, except for course 2, which needs 50 hours. This however, does not include the time spent on project work or additional discussion of case studies. If training is on all the five courses and is imparted at the rate of five hours per day in a five-day week, the entire program can be completed in six weeks.

The training hours required for each course are given below in Table 8.1.

Table 8.1: Summary of Training Courses and Number of Hours Needed

Course	No. of hours proposed for training			
	Lectures / Discussions / Demonstrations	Exercises	Visits	Total
1) Introduction to statistics The general nature and scope.	9	13	3	25
2) Introduction to statistical principles and methods.	18	24	8	50
3) Designing data collection	12	10	3	25
4) Processing, tabulation and presentation of statistics.	10	12	3	25
5) Statistical systems, coordination and standards.	9	10	6	25

Table 8-2 gives more detail on a few of the potential training topics and other activities that should lead to greatly improved data quality with minimal expenditure of time and money. It includes suggestions for training at each level and for support that could be given to improve the environment and durability of the statistical work.

Table 8.2: Examples of Training Topics and the Appropriate Level of Trainees

Training Topics	Village	District	Governorate	National
Importance of Agricultural Data	X	X	X	X
- do not let local leaders influence your estimates	X	X	X	X
Why Gather?	X	X	X	X
How used?	X	X	X	X
Importance of truth and accuracy	X	X	X	X
Free Dissemination	X	X	X	X
How it benefits farmer, village, district, governorate, national	X	X	X	X
Confidentiality for sensitive data, regulatory or tax issues	X	X	X	X
- do not give specific farmer information to leaders, especially if confidential,	X	X	X	X
operational or tax related	X	X	X	X
How to gather and record data	X		X	X
Build farmer trust and confidence in you and your services	X		X	X
What data is important	X		X	X
Observation and measurement tips	X		X	X
Data handling techniques	X		X	X
Quality control techniques	X		X	X
Probing techniques	X			
Statistical Training				
Data types	X	X	X	X
Descriptive measures of centrality and dispersion	X	X	X	X
What statistics are appropriate for each type of data	X	X	X	X
Acreage, Yield, Production, Costs of inputs, Cost of production,	X	X	X	X
family stocks, consumption, health issues	X	X	X	X
Outlier checks	X	X	X	X
Histograms, frequency distributions	X	X	X	X
Variances		X	X	X
Sampling and sample size			X	X
Sampling frame construction			X	X
Simple statistical analysis			X	X

Training Topics	Village	District	Governorate	National
More complex statistical analysis			X	X
Sample allocation			X	X
Checking techniques for verification of Village data		X		
Checking techniques for verification of District data			X	
Checking techniques for verification of Governorate data				X

8.3.3 Development of New Statistical Methods and Sampling Techniques

Most of the procedures followed in collecting data need to be further developed. EAS has already started this process by constructing a team of those who have enough experience in this field. They need to redesign questionnaires. In addition, they need to redesign the intermediate and final tables to be sent to the higher levels. Of course, the statisticians need to be trained on how to use these materials.

On the other hand, the Sampling General Directorate needs to investigate the current sampling frames and devise improvements. The current sampling techniques need to be reviewed to determine if there are any better technique to be applied. The new generation of statisticians who are or will be working in sampling activity need training to understand the theory of sampling and its applications in the field.

With regard to statistical methods, the following steps should be adopted:

Review procedures and survey methods used elsewhere in the world:

- Determine if applicable to use in Egypt,
- How could they be adapted for use in Egypt,
- Do research to develop and improve methods.

Literature search to determine if new research, not yet in use, holds potential methods that could be beneficial to Egypt's Statistical program.

Set up research to develop parameters for forecasting models.

ANNEX

Annex A contains the following:

- 1- Comparisons of the data compiled by the MVE team concerning area, yield, production, and cost of production of cotton, rice, wheat and maize at the three levels (district, governorate, and published data) for each district in Behira and Dakahlia governorates.
- 2- Comparisons of the data compiled by the MVE team concerning area, yield, production, and cost of production of cotton, wheat, maize, sorghum and fava beans at the three levels (district, governorate, and published data) for each district in Beni Suef and Assuit governorates.
- 3- Comparisons of the data compiled by the MVE team at the village level and corresponding data at both the district and governorate levels for each of the studied crops (cotton, wheat, maize and rice in Behira and Dakahlia, sorghum and fava beans in Beni Suef and Assuit governorates).
- 4- The data collected concerning the farmgate prices at various levels of the study.

Annex B contains all of the MVE team survey results concerning the following:

- 1- The agricultural extension agents' and the cooperatives managers' training, supplies, available stationery, relationship with other organizations' staff, and available data at the village level.
- 2- The main characteristics of the statisticians in the agricultural directorate at the district level and agricultural affairs department at the governorate level. This includes their qualifications, training, and procedures to be followed in collecting, tabulating, checking, and sending data to the next higher level.

Annex C contains the data collected by the MVE team concerning cost of production items for cotton, rice, wheat, maize in the studied governorates.

Annex D contains the detailed results of the field trips conducted by the MVE team to the studied governorates and the selected villages within each selected district.

Annex E contains the questionnaires used by the MVE team and guidelines for the researchers who conducted the survey about data availability and quality.

ANNEX A

ANNEX B

ANNEX C

ANNEX D

ANNEX E